

Genetic analysis of bacteriophage HK97 prohead assembly and head protein crosslinking

by

**DANJU TSO**

B. S. in Tzu Chi Medical College, Taiwan, 1998

M. S. in Institute of Biotechnology, National Cheng Kung University, Taiwan, 2000

Submitted to the Graduate Faculty of  
School of Arts & Sciences in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH  
SCHOOL OF ARTS & SCIENCES

This Genetic analysis of bacteriophage HK97 prohead assembly and head protein crosslinking  
was presented

by

Danju Tso

It was defended on

January 28, 2010

and approved by

Karen Arndt, Professor, Biological Sciences

Saleem A. Khan, Professor, Microbiology and Molecular Genetics

Craig Peebles, Professor, Biological Sciences

Anthony Schwacha, Assistant Professor, Biological Sciences

Dissertation Advisor: Roger Hendrix, Professor, Biological Sciences

Copyright © by Danju Tso

2010

# Genetic analysis of bacteriophage HK97 prohead assembly and head protein crosslinking

Danju Tso, PhD

University of Pittsburgh, 2010

*Escherichia coli* bacteriophage HK97 assembles its mature capsid from 415 copies of the major capsid protein gp5. After assembly with maturation protease, the precursor capsid called prohead undergoes proteolysis and several stages of maturation including expansion and crosslinking to form a mature capsid. During capsid assembly and maturation, gp5 proteins assemble with each other and undergo refolding and conformational changes. Thus, some of the gp5 residues must play essential roles in the local network of intra- and inter- protein interactions, prohead assembly, and the different stages of capsid maturation. Here I establish a genetic suppressor approach to investigate the functional features of gp5 residues *in vivo*. By correlating the results from biochemical, structural, and genetic studies, specific protein interactions during HK97 capsid assembly and maturation have been determined. First I report that gp5 residue V163 facilitates HK97 head protein crosslinking, possibly by providing a local hydrophobic environment to promote crosslinking. I demonstrate that mutant gp5 V163D exhibits a specific defect in the head protein crosslinking reaction. Genetic results show that Val, Leu, Thr, Ile and Cys are tolerated at gp5 residue 163, suggesting a certain limited size and its hydrophobicity are important. Second, I propose that gp5 residues 231 and 178 play central and essential roles in HK97 capsid assembly by mediating assembly of proheads from capsomers. Single substitutions at residue 231 or 178, *e.g.* D231L, block the assembly of capsomers into proheads. Interestingly, the deficiency of the mutant substitution D231L can be suppressed by substitutions K178V,

K178I, or K178N. Instead of the wild type pair of residues D231 and K178, alternative pairs, D231L/K178V, D231L/K178I, and D231L/K178N, yield viable phage HK97. Structural analysis confirms that this interaction between residues 231 and 178 is involved at the early HK97 prohead assembly from capsomers. Both cases underline that genetic suppressor studies offer useful *in vivo* data to determine the functional importance of the HK97 gp5 residues.

## TABLE OF CONTENTS

<b>PREFACE.....</b>	<b>XII</b>
<b>1.0 INTRODUCTION.....</b>	<b>1</b>
<b>1.1 BACTERIOPHAGE HK97.....</b>	<b>2</b>
<b>1.2 STUDYING THE FUNCTIONAL GP5 RESIDUES .....</b>	<b>13</b>
<b>1.3 LITERATURE REVIEW OF ISOLATING SUPPRESSORS FROM BACTERIOPHAGES AND VIRUSES .....</b>	<b>20</b>
<b>2.0 MATERIALS &amp; METHODS.....</b>	<b>31</b>
<b>2.1 MATERIALS .....</b>	<b>31</b>
<b>2.2 METHODS.....</b>	<b>37</b>
<b>3.0 GP5 RESIDUE SUBSTITUTION IN HK97 LYSOGENS BY RECOMBINEERING .....</b>	<b>51</b>
<b>3.1 INTRODUCTION .....</b>	<b>51</b>
<b>3.2 RESULTS .....</b>	<b>57</b>
<b>4.0 THE ROLE OF THE GP5 RESIDUE 163 DURING BACTERIOPHAGE HK97 HEAD PROTEIN CROSSLINKING.....</b>	<b>62</b>
<b>4.1 ABSTRACT.....</b>	<b>62</b>
<b>4.2 INTRODUCTION .....</b>	<b>64</b>
<b>4.3 RESULTS .....</b>	<b>69</b>

4.4	SUMMARY & DISCUSSION .....	84
5.0	AN ESSENTIAL INTERACTION BETWEEN RESIDUES D231 AND K178 IN HK97 PROHEAD ASSEMBLY .....	98
5.1	ABSTRACT.....	98
5.2	INTRODUCTION .....	100
5.3	RESULTS .....	103
5.4	SUMMARY & DISCUSSION .....	133
6.0	CONCLUSIONS .....	154
6.1	ISOLATING HK97 REVERTANTS FOR IDENTIFYING SUPPRESSORS IS FEASIBLE.....	155
6.2	THE LIMITATION OF THE GENETIC APPROACH .....	157
6.3	GENETICS-BASED PROTEIN ANALYSIS.....	160
	APPENDIX A.....	161
	APPENDIX B .....	164
	APPENDIX C .....	168
	APPENDIX D.....	170
	APPENDIX E .....	173
	BIBLIOGRAPHY .....	174

## LIST OF TABLES

Table 1: The recombinering bacterial strains and lysogens .....	32
Table 2: The primer pairs designed for constructing the single mutations in HK97 gp5 .....	39
Table 3: The primer pairs designed to amplify the <i>galK</i> gene for subsequent homologous recombination .....	44
Table 4: The recombination efficiency when 500nt length gp5 homologous region and 100ng, 200ng, or 400ng dsDNA were used.....	60
Table 5: The recombination efficiency when 200ng dsDNA and 110nt, 200nt, or 500nt length of the gp5 homologous region were used.....	60
Table 6: The tolerated amino acids at gp5 residue 163 found in HK97 revertants .....	81
Table 7: The amino acids Val, Thr, Cys, and Ile at residue 163 were isolated in the HK97 survivors.....	83
Table 8: The second-site suppressors of lethal mutation D231L in gp5 found in HK97 revertants .....	112
Table 9: The summary of biological activities of HK97 gp5 mutants.....	161
Table 10: The list of primers.....	173



## LIST OF FIGURES

Figure 1: The morphology of bacteriophage HK97.....	2
Figure 2: The bacteriophage HK97 capsid assembly and maturation .....	4
Figure 3: The major domains of the bacteriophage HK97 gp5 .....	7
Figure 4: The K169-N356 crosslinking on bacteriophage HK97 Head II structure.....	9
Figure 5: The amino acid sequence of the HK97 gp5 .....	16
Figure 6: The life cycle of the bacteriophage HK97.....	17
Figure 7: The outline of the genetic approach .....	19
Figure 8: The genetic map of the plasmid pV0.....	33
Figure 9: The strategy to introduce quasi-random amino acids at gp5 residue 163 .....	41
Figure 10: The phage complementation assay.....	42
Figure 11: The lambda Red proteins, Exo and Bet, provide homologous recombination in <i>E. coli</i> .....	53
Figure 12: The genetic map of the pRED plasmid pKD46 (Datsenko and Wanner, 2000; Guzman et al., 1995) .....	54
Figure 13: The recombineering-based <i>galK</i> positive/negative selection .....	55
Figure 14: The amplified products from the colonies grown on DOG plate.....	57

Figure 15: The amplified product from the lysogens, with or without a <i>galK</i> gene in the prophage HK97.....	58
Figure 16: The proposed chemical mechanism of the HK97 capsid crosslinking.....	64
Figure 17: The HK97 head crosslink site .....	65
Figure 18: The partial sequence alignment of the major capsid proteins from the dsDNA phages HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B .....	67
Figure 19: The close-up view of the HK97 head crosslink site.....	68
Figure 20: The mutant gp5 V163D cannot complement the gene 5 <sup>-</sup> amber phage.....	70
Figure 21: Prohead II V163D can be purified <i>in vitro</i> .....	72
Figure 22: Prohead II V163D can expand <i>in vitro</i> .....	74
Figure 23: Prohead II V163D cannot fully crosslink <i>in vitro</i> (treatment: SDS sample buffer)....	75
Figure 24: Prohead II V163D cannot fully crosslink <i>in vitro</i> (treatment: high salt/ low pH buffer) .....	77
Figure 25: Prohead II V163D cannot fully crosslink <i>in vitro</i> (treatment: DMF buffer).....	79
Figure 26: The tolerated amino acids at gp5 residue 163 in the isolated HK97 survivors .....	86
Figure 27: The HK97 crosslink active site .....	96
Figure 28: The T loop, an unusual sub-structure in the gp5 P domain.....	102
Figure 29: The gp5 mutants D231A and $\Delta$ loop cannot be cleaved by the HK97 protease gp4	104
Figure 30: The gp5 single mutants D231L, D231A, K178E, and K178L assemble capsomers but do not assemble proheads .....	106
Figure 31: The double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N can complement the gene 5 <sup>-</sup> amber phage.....	108

Figure 32: The gp5 single mutants D231L, K178L, and K178E cannot be cleaved by the HK97 protease gp4 .....	110
Figure 33: The single subunit gp5 from the HK97 Prohead II (a) and the HK97 Head II (b) structures .....	116
Figure 34: The interactions of residues D231 and K178 between HK97 Prohead II capsomers	121
Figure 35: The double mutant Prohead Is D231L/K178N, D231L/K178I, and D231L/K178V are not detected .....	124
Figure 36: The double mutant Prohead IIs D231L/K178N, D231L/K178I, and D231L/K178V are detectable .....	126
Figure 37: Proteolysis of three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N .....	128
Figure 38: The single gp5 mutants K178E and K178L do not produce functional gp5 .....	130
Figure 39: The residues D231 and K178 on the HK97 Prohead II (a and b) and the HK97 Head II (c and d) structures .....	138
Figure 40: The relationship of the residues D231 and K178 on the HK97 Head II .....	141
Figure 41: The sequence pairs of the gp5 residues 231 and 178 in the wild type HK97 and the three revertants .....	143
Figure 42: The partial sequence alignment of the major capsid proteins from the dsDNA phages HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B .....	151
Figure 43: The major capsid protein sequence alignment of the dsDNA bacteriophage HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B .....	167
Figure 44: The HK97 head crosslink site .....	169
Figure 45: The HK97 prohead assembly .....	172

## **PREFACE**

It is my honor to be a member of Dr. Roger Hendrix's lab. Dr. Hendrix is a great mentor and has always been very patient and insightful when he discussed science with me. He serves as my role model as a professor and thesis advisor and I have learned a lot from him. I would like to thank my thesis committee members: Prof. Karen Arndt, Prof. Saleem Khan, Prof. Craig Peebles, and Prof. Anthony Schwacha. I am grateful to have Prof. Arndt, a successful female scientist, in my committee. I appreciate that Dr. Khan and Dr. Peebles schedules me into his busy calendar. I will thank Dr. Schwacha provides lots of suggestions.

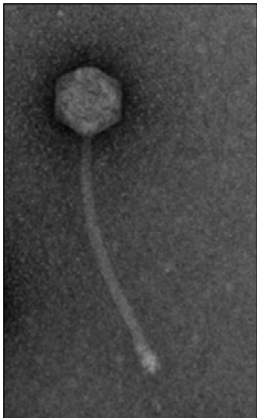
I would like to thank people in Dr. Hendrix's lab and all my friends in Pittsburgh. Dr. Robert Duda provides lots of information about HK97 structural analysis. Dr. Craig Peebles and Dr. Susan Godfrey were very helpful at providing ideas for my scientific works. Brian always keeps the lab under the best working condition. My wonderful labmates: Dr. Welkin Pope, Ching-Chung, Dan in room 342, and Brian, Robert, and Jianfei are nice partners and good friends. All of the fun time together will be my best memories. A specially thank to Dr. Don Court for providing the recombineering strains and plasmids. Studying in University of Pittsburgh, USA, is a great adventure to me. I enjoyed being blessed to have such an opportunity. I want to share the pleasure of finishing desertation with my mother, my grand parents, and my husband, Dr. David Liu.

## 1.0 INTRODUCTION

Bacteriophages (phages) are thought to be the most abundant organisms on the earth and are pervasive in our environment. They are found in essentially all environments where their bacterial hosts are found, such as soil, ocean, even in the human digestion system (Chibani-Chennoufi et al., 2004; Gill et al., 2006; Rusch et al., 2007; Zhang et al., 2006). For nearly 100 years, studies of bacteriophages have contributed a stream of fundamental knowledge in the areas of genetics, molecular biology, and structural biology. Among all bacteriophages, the tailed dsDNA bacteriophages are the most plentiful, accounting for ~95% of the isolates described in the literature. They are exceptionally well-studied in relation to viral gene regulation, virus-host interactions, and capsid and tail assembly (Ackermann, 1992; Chibani-Chennoufi et al., 2004). Bacteriophage HK97 (Figure 1), using dsDNA as its genetic material, is a well-developed and successful model for addressing questions about dsDNA bacteriophage capsid assembly (Duda et al., 1995b; Hendrix and Duda, 1998; Popa et al., 1991a).

## 1.1 BACTERIOPHAGE HK97

Bacteriophage HK97 is a temperate phage of *Escherichia coli* (*E. coli*), which belongs to the family *Siphoviridae* (Ackermann, 1992; Ackermann and Kropinski, 2007). It was isolated in Hong Kong (Dhillon et al., 1980). The morphology of the HK97 virion includes an icosahedral head (capsid) with a long flexible, non-contractile tail (Figure 1). Our laboratory has studied the HK97 capsid for many years. We discovered that HK97 has an unusual and interesting course of capsid assembly and maturation (Conway et al., 1995; Duda et al., 1995a; Duda, Martincic, and Hendrix, 1995; Duda et al., 1995b; Popa et al., 1991a).



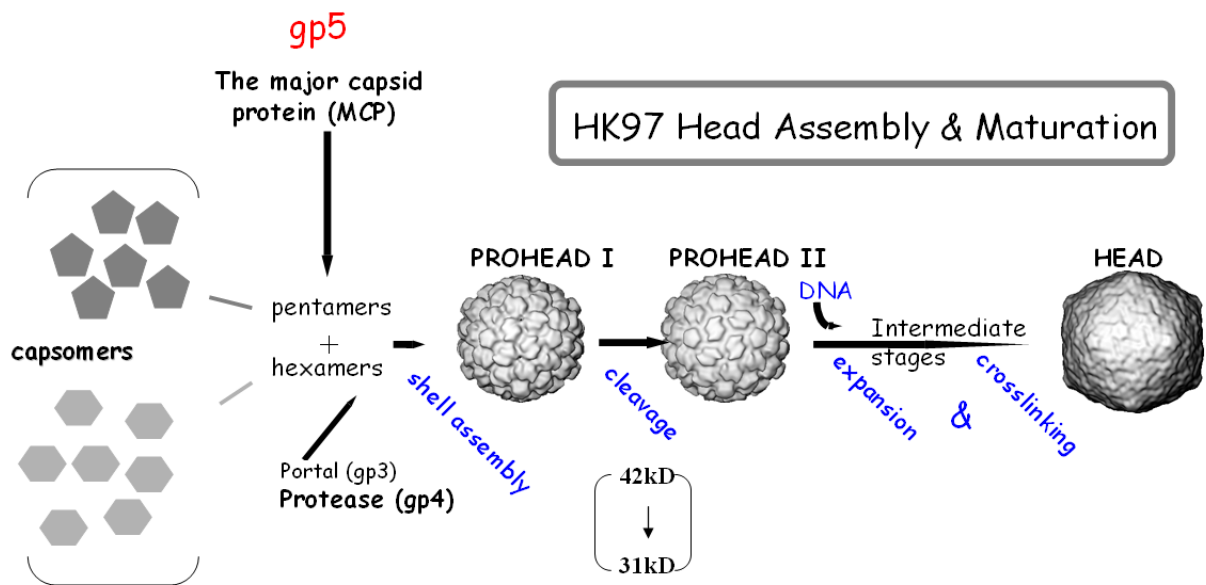
**Figure 1: The morphology of bacteriophage HK97**

The bacteriophage HK97 particle was examined by transmission electron microscopy. The sample was prepared from a crude lysate and the particles were stained with uranyl acetate. HK97 has an icosahedral head and a long flexible tail. The micrograph was taken by Danju Tso.

### **Assembly of bacteriophage HK97 capsid**

The major capsid protein of HK97 is encoded by gene 5 (Juhala et al., 2000), and it is usually called gp5 (gp=gene product). Over-expression of gp5 from a plasmid in *E. coli* results in efficient assembly of icosahedrally symmetric shells, termed Prohead I (Figure 2) (Duda, Martincic, and Hendrix, 1995; Xie and Hendrix, 1995). Prohead I contains 420 copies of the major capsid protein gp5 arranged in a shell with triangulation number 7 (T=7) (Hendrix and Duda, 1998; Wikoff et al., 1998). The 420 subunits of Prohead I are arranged in 60 hexamers and 12 pentamers (Figure 2) (Hendrix and Duda, 1998; Wikoff et al., 1999).

HK97 Prohead I can be dissociated into a mixture of pentamers and hexamers, collectively known as capsomers, when Prohead I is treated with high salt, *e.g.* 2M KCl (Xie and Hendrix, 1995). The dissociated capsomers can re-assemble into Prohead I when the capsomers are incubated with polyethylene glycol (PEG) and certain ionic solutions, *e.g.* 4% PEG and 10mM CaCl<sub>2</sub> (Xie and Hendrix, 1995). It was proposed that HK97 gp5 can first assemble into two different capsomers: hexamers and pentamers, as the initial step of assembly, and these capsomers assemble into Prohead I (Figure 2). However, the detailed mechanism by how gp5's assemble into capsomers, how capsomers assemble into proheads, and how that assembly is regulated to produce only the capsid of the correct size, remains unknown.



**Figure 2: The bacteriophage HK97 capsid assembly and maturation**

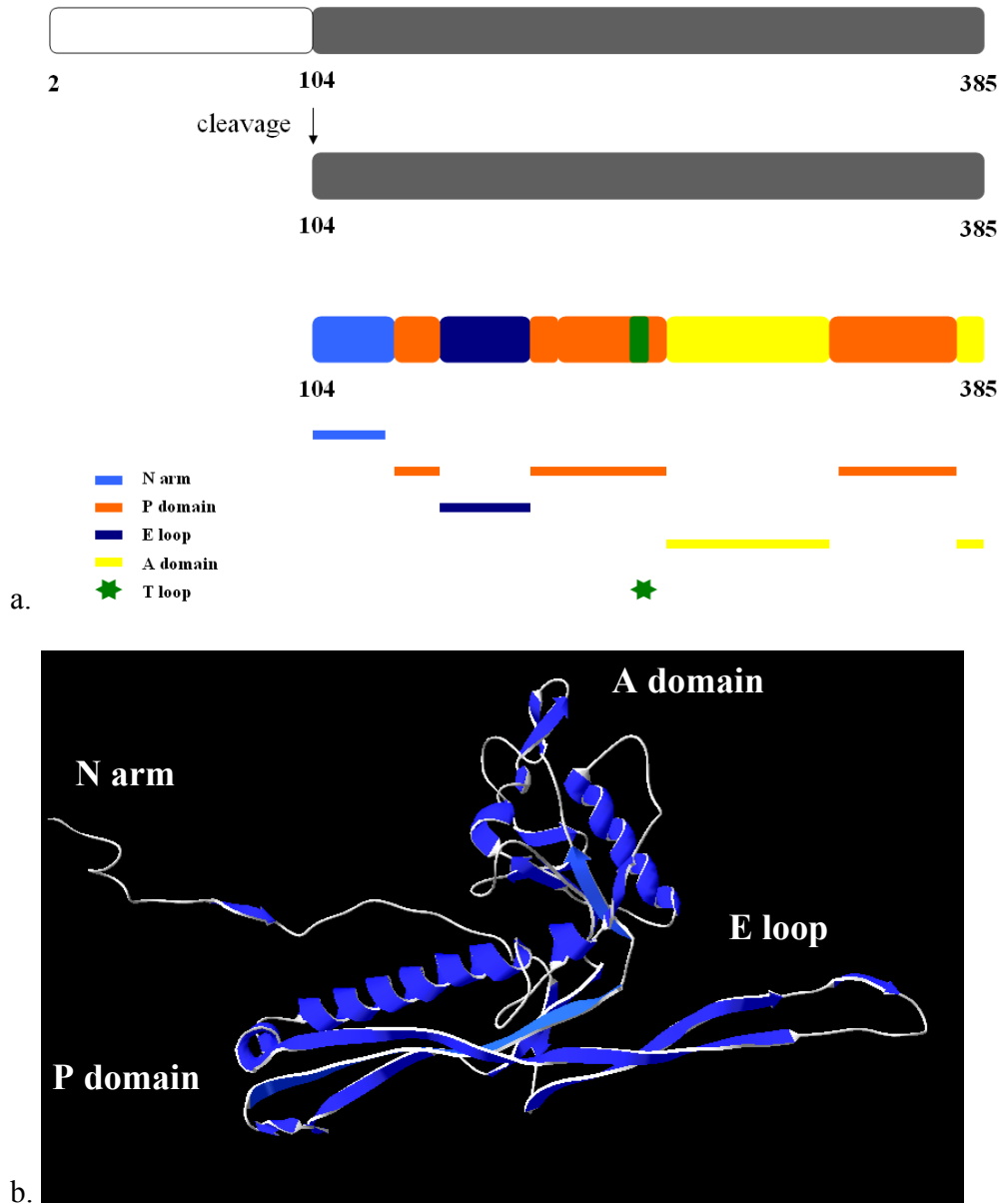
HK97 major capsid protein gp5 assembles into two capsomers: pentamers and hexamers. These capsomers self-assemble into a large shell, called Prohead I. Proteolytic cleavage triggers Prohead II formation in which the N-terminal part of gp5 is cleaved by the protease gp4, decreasing the size of gp5 from 42kD to 31kD. Prohead II expansion is triggered by DNA packaging and entails several stages of conformational changes, remodeling the HK97 capsid. Covalent crosslinking among gp5 subunits takes place progressively during capsid expansion and secures the shell to form a mature HK97 capsid. Portal protein gp3 is not essential for Prohead I assembly *in vitro*, but portal function is required for DNA packaging and viable phage production.



In a normal HK97 infection, there are two additional phage-encoded proteins with roles in establishing the mature structure of the capsid. The first of these is the 47kD product of gene 3, gp3, called portal protein (Juhala et al., 2000). The portal is a ring-shaped dodecamer of gp3, which is present once in the capsid shell, replacing one of the 12 icosahedral vertices of the shell in place of one of the gp5 pentamers that would otherwise be there (Hendrix and Duda, 1998). Portal has crucial roles in assembly and function of the capsid, as the entry and exit port for the DNA, as a component of the DNA packaging pump, and as the attachment site for the tail (Fokine et al., 2005; Johnson and Chiu, 2007; Leiman et al., 2003; Petrov and Harvey, 2008; Rao and Feiss, 2008). However, portal appears not to have any significant roles in maturation of the capsid, and many of the studies on that aspect of capsid assembly have been done with HK97 structure lacking a portal.

Protein gp4 (~25kD) is responsible for proteolysis of gp5 (Conway et al., 1995; Duda, Martincic, and Hendrix, 1995). This proteolytic cleavage decreases the size of gp5 from 42kD to 31kD (Conway et al., 1995; Duda, Martincic, and Hendrix, 1995). After 50-100 copies of the protease co-assemble with gp5, the protease gp4 cleaves off 102 residues (residue 2-103) from the gp5 amino terminus—a region called Delta domain—producing Prohead II (Figure 2, 3) (Conway et al., 1995; Duda, Martincic, and Hendrix, 1995; Juhala et al., 2000; Wikoff et al., 2003). The Delta domain can be observed inside Prohead I in the cryo-EM images but is not seen in the cryo-EM images of Prohead II, consistent with its physical absence from Prohead II, as indicated by a measured mass difference between Prohead I and Prohead II (Conway et al., 1995). HK97 has no separate scaffolding protein to assist capsid assembly. Delta domain is located at the 5'-end of gene 5, which is similar to the location of the scaffolding protein genes in many other bacteriophages. It is believed that Delta domain functions as an assembly chaperone

or scaffold since HK97 can assemble into a large structure—prohead—without any separate scaffolding protein. The protease gp4 may undergo self-degradation and leave the capsid after proteolysis, since no gp4 remains on the mature capsid (Duda et al., 1995b; Popa et al., 1991a).



**Figure 3: The major domains of the bacteriophage HK97 gp5**

a. HK97 gp5 Delta domain is cleaved by the protease gp4 during capsid maturation. Proteolysis removes the N-terminal part of gp5 from residue 2 to 103. The distinct gp5 domains are N arm (for N-terminal), P domain (for Peripheral domain), E loop (for Extended loop), A domain (for Axial domain) and a sub-structure called T loop .b. The three-dimensional structure of a single gp5 from HK97 mature capsid, Head II. The gp5 is shown in color blue with the white string.

Proteolysis is followed by several additional stages of capsid maturation (Conway et al., 2001; Duda et al., 1995a; Duda et al., 1995b; Gan et al., 2004; Gan et al., 2006; Hendrix and Duda, 1998; Lata et al., 2000; Lee et al., 2004; Rader, Vlad, and Bahar, 2005; Wikoff et al., 2006). There are two significant transitions. One, HK97 Prohead II expands its size, and the HK97 shell wall becomes thinner (Duda et al., 1995a; Lata et al., 2000). Prohead shape is changed from irregular and round to angular and icosahedral (Figure 2) (Lata et al., 2000). This conformational change is a modification of the size and the shape of the capsid, and can be induced *in vitro* (Duda, Martincic, and Hendrix, 1995). HK97 Prohead II expands when Prohead II is treated with certain agents, *e.g.* Sodium Dodecyl Sulfate (SDS) sample buffer, high salt/low pH buffer, and dimethyl formamide (DMF) buffer (Dierkes et al., 2009; Duda et al., 1995a). The results of the expansion can be observed as a mobility shift on an agarose gel since the expansion changes the surface charge and the size of HK97 capsid (Duda et al., 1995a; Lata et al., 2000). Second, HK97 has no decoration protein to stabilize the viral shell, but it does form covalent crosslinks between gp5 subunits (Popa et al., 1991b). The side chain of Lys169 (K169) on one gp5 subunit forms a covalent isopeptide bond with the side chain of Asn356 (N356) on the neighboring gp5 (Figure 4) (Duda et al., 1995a; Wikoff et al., 2000). There are 420 crosslink sites on a HK97 mature capsid. It was shown that this irreversible chemical reaction is catalyzed by Glu363 (E363), which is located on a third gp5 subunit and may serve as a proton acceptor (Dierkes et al., 2009). The crystal structure of HK97 mutant capsid K169Y has been determined and named Head I. The HK97 mature capsid structure (wild type) was determined and called Head II (Gan, 2006; Helgstrand, 2003).



Crosslinking of HK97 Prohead II can be induced *in vitro* (Duda et al., 1995a). Incubating Prohead II with solutions containing detergents or denaturants, *e.g.* urea or SDS, or exposing the Prohead II to low pH buffer (pH~4), and neutralizing after exposure (pH~7), causes Prohead II to form crosslinks (Figure 4a)(Duda et al., 1995a). The crosslinking is not caused directly by these treatments; rather these treatments trigger capsid expansion, and one of the consequences of the conformational changes of expansion is assembly of the catalytic sites where crosslinking takes place (Gan et al., 2004; Ross et al., 2005).

After treatments, Prohead II undergoes the formation of different crosslinking intermediates, which can be monitored on an SDS Polyacrylamide Gel Electrophoresis (SDS PAGE) (Figure 4a). Major intermediates are determined to be monomer, dimer, trimer, tetramer, linear pentamer, linear hexamer, circular pentamer, and circular hexamer (Figure 4a). With longer exposure to the solutions, monomers thru linear hexamers disappear eventually. Circular pentamers and hexamers show up after few minute incubation (Figure 4a). A large crosslinking protein network (called chainmail), which encompasses essentially the entire capsid, remains in the loading wells, since their structures are too big to enter an SDS PAGE (Figure 4a) (Duda, 1998; Ross et al., 2005; Wikoff et al., 2000). The complete HK97 mature capsid is expanded and crosslinked (Figure 2). The protein chainmail, which connects the capsid subunits, is a result of the adjacent protein rings being catenated into a large network (Figure 4c and 4d) (Duda, 1998; Wikoff et al., 2000).

### **The bacteriophage HK97 head protein crosslinking is unusual among phages**

Prohead expansion is a universal change among dsDNA bacteriophages. It remodels HK97 capsid in such a way that the interior volume of the shell doubles, and the shape of the capsid changes (Conway et al., 1995; Lata et al., 2000). However, the irreversible covalent crosslink, which increases the capsid stability, is not commonly found in dsDNA bacteriophages. More commonly, dsDNA bacteriophages, *e.g.* lambda and T4 do not have the crosslinking on their capsids but have decoration proteins binding to their capsid for additional stability (Fokine et al., 2004; Fokine et al., 2005; Ishii and Yanagida, 1975; Ishii and Yanagida, 1977; Iwasaki et al., 2000; Lander et al., 2008; Qin et al., 2009; Sternberg and Weisberg, 1977; Steven et al., 1992; Yang, Maluf, and Catalano, 2008).

Although the HK97 head protein crosslinking is unusual, HK97 is not the only bacteriophage that does crosslink. *Pseudomonas aeruginosa* phage D3 major head protein gp6 is a homolog of HK97 gp5 (Gilakjan and Kropinski, 1999; Kropinski, 2000). It shows 43% identity of protein sequence to HK97 gp5 (Gilakjan and Kropinski, 1999). It was suggested that phage D3 mimics the capsid maturation of HK97, undergoing similar conformational transitions, *i.e.* proteolysis and crosslinking (Gilakjan and Kropinski, 1999). Phage D3 gp6 contains an N-terminal domain similar to HK97 Delta domain, and the D3 N-terminal domain is cleaved from the rest of the protein during maturation (Gilakjan and Kropinski, 1999). Besides, phage D3 capsid does crosslink and the crosslinking residues, Lys178 and Asn363, corresponding to K169 and N356 in HK97, are conserved and can form a covalent Lys-Asn isopeptide bond (the K-N isopeptide bond) (Gilakjan and Kropinski, 1999). Some of the mycobacteriophages, *e.g.* L5, also carry out capsid crosslinking (Hatfull and Sarkis, 1993).

Isopeptide bonds between lysine and glutamine are found in numerous examples across the biosphere (Lee et al., 2001; Mosesson et al., 1995). The HK97 capsid crosslink differs from these examples in two significant ways: first, it links lysine to asparagine rather than to glutamine, and second, its formation is not catalyzed by an enzyme (Lorand and Graham, 2003; Murthy et al., 1991; Murthy et al., 2000; Tatsukawa et al., 2009). Recently, another crosslink with these characteristics has been reported in the pili of the Gram-Positive bacterium *Streptococcus pyogenes* (Kang and Baker, 2009; Kang et al., 2007). It was reported that two intramolecular K-N crosslinks form inside the backbone pilin protein spy0128. These bonds link Lys36 to Asn168 in the N domain, and Lys179 to Asn303 in the C domain of spy0128 (Kang et al., 2007). It is interesting to note that a Glu residue is in the near proximity of each bond, Glu117 in the N domain and Glu258 in the C domain (Kang et al., 2007). The geometric relationships among the three crosslink-related residues, Lys, Asn, and Glu in the *S. pyogenes* pilin protein are very similar to the corresponding geometry at the HK97 crosslink site, suggesting a similar chemical mechanism for crosslinking formation (Dierkes et al., 2009). It seems likely that more examples of the K-N crosslink will be discovered in the future.



## 1.2 STUDYING THE FUNCTIONAL GP5 RESIDUES

HK97 major capsid protein gp5 undergoes a remarkable series of changes during capsid assembly and capsid maturation. At the beginning of assembly, all of the single subunits—gp5's—assemble with each other to form two different capsomers: pentamers and hexamers. The capsomers then assemble into a large shell structure, the prohead. During the maturation stages of expansion and crosslinking, each gp5 undergoes significant conformational changes and forms covalent crosslinks to its neighbors. Thus, the protein gp5 has to change its shape, properties, and functions as it moves through each step of the maturation pathway. These observations indicate that gp5 is an extremely flexible protein and has to interact with other gp5's in a precise, stable, but dynamic way. Structural studies have revealed a picture at moderately high resolution of the morphological changes of the HK97 capsid as it matures. For instance, the structure of a single protein gp5 on HK97 Head II differs from that on HK97 Prohead II (Conway et al., 2001; Gertsman et al., 2009; Helgstrand et al., 2003), indicating that every single subunit undergoes conformational changes during capsid expansion. However, the morphological transitions demonstrated by the structural visualization do not reveal the functional significance of the interactions of different parts of the protein with each other or with the neighboring proteins in the structure.

### **The previous mutational analysis of gp5 residues**

The HK97 mature phage particle contains 415 copies of the major capsid protein, and the major capsid protein gp5, containing 384 residues, folds with the assistance of the *E. coli* GroE chaperonins into a stable three-dimensional structure (Figure 3) (Helgstrand et al., 2003; Wikoff et al., 2000). After proteolysis, the protease gp4 cleaves off the Delta domain of gp5, composed of 102 residues (Figure 3) (Duda et al., 1995a). Based on the HK97 Head II structure, four major distinct domains of gp5 have been defined: the N arm (=N terminal), the A domain (=Axial domain), the P domain (=Peripheral domain), and the E loop (=Extended loop) (Figure 3) (Helgstrand et al., 2003). In the past few years, the functions of certain gp5 residues have been analyzed in our laboratory by mutagenesis. Some of the single substitutions in gp5 residues can be tolerated without affecting the function of the gp5, but changes some other residues can be tolerated minimally, or not at all (Figure 5, Appendix A). The results of the mutational studies of gp5 provide a better understanding of HK97 capsid.

Residues K169, N356, and E363 mentioned above have been shown to be vital for the formation of the HK97 covalent K-N crosslink bond (Figure 4). For example, substitution of Lys at residue 169 with Tyr (K169Y) or Asn (K169N), or of Asn at residue 356 with Asp (N356D) prevent crosslinking (Duda, Martincic, and Hendrix, 1995). Single mutation K169Y or K169N in the protein gp5 specially causes a failure of HK97 protein crosslinking since gp5 can be cleaved and expand. The protein gp5 with single mutation N356D can assemble capsomers but cannot assemble proheads. In addition, the substitutions of Glu at residue 363 with Asp, Ala, His, Lys, Asn, Gln, or Tyr, also block the HK97 capsid crosslinking (Dierkes et al., 2009). All of the above data indicate that a single substitution at each residue 169, 356, or 363 within gp5 is sufficient to disrupt HK97 head protein crosslinking. However, some residues in gp5 are not

needed in their original form. The substitution of either Asp108 or Aps110 to Ala did not cause a significant defect in the functionality of the gp5 (Figure 20 and data not shown), suggesting that residue 108 or 110 are not important for the network interactions of protein gp5's on the HK97 capsid.

The mutational analysis of gp5 residue 219 reveals information about capsomers. The mutant Prohead I E219K is morphologically indistinguishable from wild type Prohead I (Li et al., 2005). However, unlike the wild type, no pentamer was found after Prohead I E219K was dissociated (Li et al., 2005). Interestingly, mutant E219K hexamers reassemble into a capsid-like structure without pentamers, called “Whiffleball,” and the absence of the pentamer in “Whiffleball” can be fixed by adding wild type pentamers, implying that wild type pentamers can cooperate with mutant E219K hexamers (Li et al., 2005). Cryo-EM confirmed that “Whiffleball” is indistinguishable from wild type Prohead I except for the lack of pentamers (Li et al., 2005), suggesting that the residue E219 may play an essential role in assembling pentamers, as well as assembling pentamer with hexamers.

Some single substitutions in the gp5 residues cannot complement the gene 5' amber phage (Appendix A), indicating that the mutated residues may play critical roles in gp5 folding, prohead assembly, and capsid maturation. In my studies, I focus on investigating the critically functional roles of gp5 residues 163 and 231, located in the E loop and the P domain, respectively.

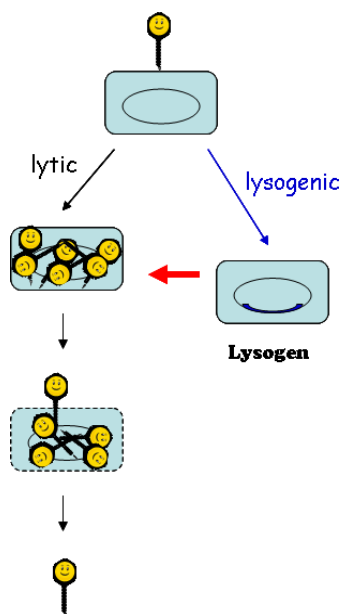
1	MSELALIQKA	IEESQQKMTQ	LFDAQKAEIE	STGQVSKQLQ	SDLMKVQEEL
51	TKSGTKLFDL	EQKLASGAEN	PGEKKSFSER	AAEELIKSWD	GKQGTFGAKT
101	FNKSLGSDAD	SAGSLIQPMQ	IPGIIMPGLR	RLTIRDLLAQ	GRTSSNALEY
151	VREEVFTNNA	DVVAEKALKP	ESDITFSKQT	ANVKTIAHWV	QASRQVMDDA
201	PMLQSYINNR	LMYGLALKEE	GQLLNGNGTG	DNLEGLNKVA	TAYDTSLNAT
251	GDTRADIIAH	AIYQVTESEF	SASGIVLNPR	DWNHIALKLD	NEGRYIFGGP
301	QAFTSNIMWG	LPVVPTKAQA	AGTFTVGGFD	MASQVWDRMD	ATVEVSREDR
351	DNFVKNMLTI	LCEERLALAH	VRPTAIKGT	FSSGS	

**Figure 5: The amino acid sequence of the HK97 gp5**

HK97 major capsid protein gp5 contains 385 amino acids. The region marked in color gray is the Delta domain. The residues in red have been analyzed by mutational analysis in this and earlier studies, and their results are shown in Appendix A.

### Isolation of suppressors of lethal mutations in the bacteriophage HK97 gp5

Bacteriophage HK97 is a temperate phage, which can undergo its lytic cycle to produce progeny virions. It can also integrate its genome into a host to develop a lysogen, called lysogenic cycle (Figure 6). The integrated genome is called a prophage, and the host containing the prophage is called a lysogen. Because the prophage is replicated by the host replication machinery, it is possible to introduce and propagate a lethal mutation in the HK97 prophage in a lysogen. During lytic growth, the lethal mutation in gp5 may cause defects in the gp5 folding, prohead assembly and capsid maturation. The constructed HK97 lysogen can be induced for isolating revertants, which may contain suppressors. Thus, the various suppressors can be identified and the network of interactions in HK97 capsid can be illuminated. HK97 needs only one kind of protein, gp5, to build the entire protein shell. It is not unexpected that only intragenic suppressors have been isolated. These intragenic suppressors can provide important intra- (within a subunit gp5) or inter-(between subunits) molecular interactions to rescue the HK97 capsid assembly and maturation.

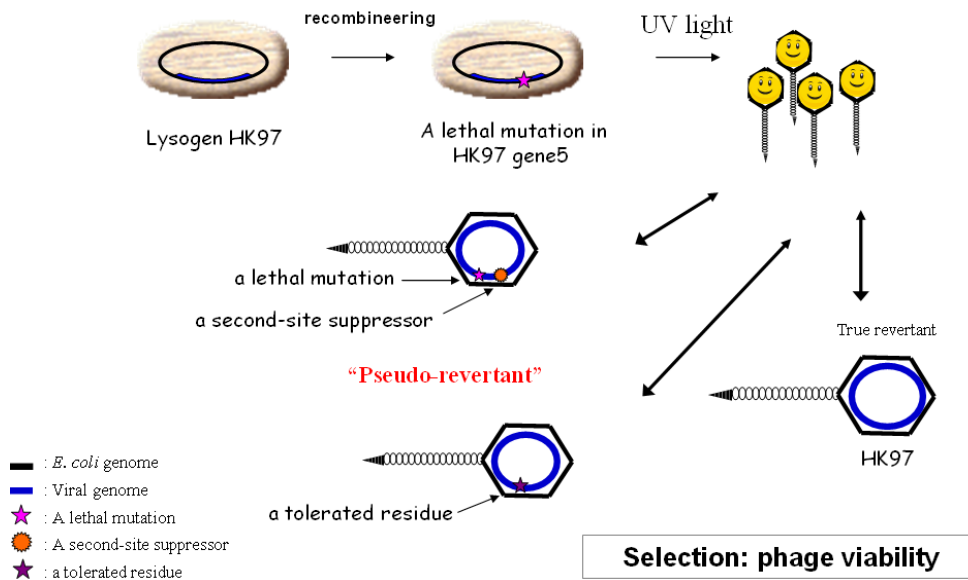


**Figure 6: The life cycle of the bacteriophage HK97**

## **The thesis goals**

The goal of this study is to obtain functional information about the roles of specific amino acids in the HK97 major capsid protein, gp5, e.g local network interaction. This information can be correlated with other available information about the HK97 system, particularly the results from structural and biochemical studies, to give a more complete picture of the capsid assembly and maturation, and gp5 residues' functions.

A genetic approach was used in my study: a lethal mutation was generated in the gene 5 of a HK97 prophage (Figure 7). The mutant lysogen should not produce plaque-forming phage particles, since it does not encode a functional gp5. However, they are able to be induced in our laboratory to obtain revertants for re-gaining phage viability. The surviving revertants which contain suppressors can be isolated. Revertants that arose from the mutant HK97 lysogen are either “pseudo-revertants” or “true revertants” (Figure 7). “Pseudo-revertants” contain either a non-wild type replacement of the original mutant residue or a second-site suppressor (Figure 7). The first pseudo-revertant has a functional residue that is not the original residue. The second-site suppressor is any residue that appears at a position other than the original one and can suppress the lethal mutation. These suppressors can provide *in vivo* information of the subunit interactions on HK97 capsid. “True revertants” represent a reversal of the lethal mutation to the mutation to restore the original residue in gp5 (Figure 7). Unlike pseudo-revertants, they cannot offer any genetic information and may be the most common of the isolated revertants.



**Figure 7: The outline of the genetic approach**

A lethal mutation is introduced into lysogen HK97 for replacing one gp5 residue by using the technique, recombineering. Recombineering will be described in Chapter 3. The mutant lysogen HK97, which contains a lethal mutation in HK97 gp5, is induced by UV light for producing phage particles. Either pseudo-revertant or true revertant will be isolated based on the phage viability selection. Pseudo-revertant carries either a non-wild type replacement of the lethal mutation (a tolerated residue), or the lethal mutation and second site suppressors to suppress the lethal mutation.

### **1.3 LITERATURE REVIEW OF ISOLATING SUPPRESSORS FROM BACTERIOPHAGES AND VIRUSES**

Isolation of second-site suppressors is a useful technique to study the structural genes in bacteriophages and viruses, which have been found in the history of studying phage P22, Lambda,  $\Phi$ X174, f1, and some eukaryotic viruses (see below). Second-site suppressors may be isolated within the mutated gene (intragenic suppression) or outside of the mutated gene (intergenic suppression). This section focuses on discussing the suppressors isolated from the structural genes of phages and viruses.



### **The intragenic suppressors isolated from *Salmonella* phage P22**

*Salmonella typhimurium* dsDNA bacteriophage P22 is a model for studying phage capsid and tail assembly. The intragenic suppressors isolated from the phage P22 coat protein and the tailspike protein have been extensively investigated (Aramli and Teschke, 1999; Bazinet, Villafane, and King, 1990; Beissinger et al., 1995; Fane and King, 1991; Fane et al., 1991; Fong, Doyle, and Teschke, 1997; Gordon and King, 1994; Lee, Koh, and Yu, 1991; Parent, Ranaghan, and Teschke, 2004; Parent, Suhanovsky, and Teschke, 2007).

Intragenic suppressors, D163G, T166I, and F170L, have been identified in the P22 coat protein which globally suppress the temperature sensitive (ts) folding defects (Aramli and Teschke, 1999). Multiple isolations from different strains which carry different ts mutations suggested that they are global suppressors that can reverse several ts mutations (Aramli and Teschke, 1999). For instance, the single substitution T166I can globally suppress single mutations, S223F and F353L, in the P22 coat protein (Aramli and Teschke, 1999; Parent, Ranaghan, and Teschke, 2004; Parent, Suhanovsky, and Teschke, 2007). By studying the ts mutants S223F and F353L, and the double mutants S223F/T166I and F353L/T166I, the interaction of the coat protein and chaperone network was determined (Parent, Ranaghan, and Teschke, 2004). In comparison to the wild type coat protein, the suppressor mutants S223F/T166I and F353L/T166I induce more chaperones GroEL and GroES and produce more procapsid (Parent, Ranaghan, and Teschke, 2004). Interestingly, expression of the double or triple suppressor mutants coat proteins, D163G/T166I, D163G/F170I, T166I/F170I, and D163G/T166I/F170I, results in cold-sensitive (cs) phenotypes (Parent, Ranaghan, and Teschke, 2004). Furthermore, these single, double, and triple suppressor mutants assemble more aberrant structures than the wild type, and double and triple suppressor mutants did not undergo the

proper expansion *in vitro* (Parent, Ranaghan, and Teschke, 2004). It indicates that these suppressor mutants can cause defective phenotypes on the phage P22 and they may rescue the ts phenotype of P22 by reversing or changing the properties of the P22 coat protein. After analyzing the localization of the above three global suppressors, it was found that they were located in the flexible hinge region, comprising the residues 157-207, in the P22 coat protein (Parent, Suhanovsky, and Teschke, 2007). So the above suppression may rescue the coat protein through the residue interactions related to the hinge region in P22 coat protein.

Other intragenic suppressors isolated from the phage P22 were identified for the ts mutations of the tailspike protein. P22 tailspike proteins non-covalently bind to mature phage P22 capsids, and the function of the P22 tailspike protein is the recognition of the phage P22 receptor (Berget and Poteete, 1980). The ts folding mutations were determined in the P22 tailspike protein, encoded by gene 9, and the second-site suppressors were isolated from these ts mutants (Bazinet, Villafane, and King, 1990; Fane and King, 1991; Fane et al., 1991; Mitraki et al., 1993; Villafane, Fleming, and Haase-Pettingell, 1994). The ts mutations in the P22 tailspike protein reduce the refolding of the tailspike trimer and increase the protein aggregation while the intragenic second-site suppressors recover the above defect by enhancing the refolding and inhibiting the protein aggregation (Mitraki et al., 1993). Specifically, two global suppressors V331A and A334V can complement the defects that were caused by some of the single substitutions in the tailspike protein (Fane et al., 1991; Villafane, Fleming, and Haase-Pettingell, 1994). These global suppressors were independently identified from multiple strains and can correct the tailspike protein ts defects (Fane et al., 1991; Villafane, Fleming, and Haase-Pettingell, 1994). Interestingly, residues 331 and 334 were found located close together, and their ts parental mutations were mapped to the central region of the P22 tailspike protein. This

suggests that there is a local interaction between the central region and the region around residues 331 and 334. Since the ts mutants affect the early stage of the folding intermediates, not the stability or functions of the native protein, the above local interaction should be important for the folding intermediates of the P22 tailspike protein.

Single mutation hmH3034 in the phage P22 gene 9 (called D100N in this report) causes the tailspike trimers to partially unfold, and also decreases the affinity of the tailspike to capsid about 100-500 fold (Maurides, Schwarz, and Berget, 1990). Intragenic suppressors R13H, R13L, and R13S, which substitute His, Leu, and Ser for the wild type amino acid Arg, were isolated from the mutant D100N individually, indicating that these second-site suppressions in the P22 restore the normal function of P22 tailspike (Maurides, Schwarz, and Berget, 1990). Double mutants D100N/ R13H, D100N/R13L, D100N/R13S and the single suppressor mutants R13H, R13L, and R13S form fully functional tailspike proteins with assembly activities indistinguishable from wild type. Single substitutions R13H, R13L, and R13S did not cause any phenotypic defect (Maurides, Schwarz, and Berget, 1990). These genetic data suggested that residues D100 and R13 form a salt bridge which stabilizes the amino terminus of the tailspike protein (Maurides, Schwarz, and Berget, 1990).

### **The intragenic suppressors isolated from murine coronavirus virus, yellow fever virus, and human immunodeficiency type 1 virus**

The second-site suppressors R654H/E1035D of the substitutions Q159L/ H716D were isolated from the murine coronavirus strain A59 spike protein (Navas-Martin, Hingley, and Weiss, 2005). The revertant containing Q159L/R654D/H716D/E1035D in the spike protein was isolated from the mutant Q159L/H716D by a serial passage in mice until hepatitis was observed (Navas-Martin, Hingley, and Weiss, 2005). One of the original substitutions Q159L located in the putative receptor binding domain (RBD) is associated with the loss of the ability to induce hepatitis, and the other substitution H716D found within the cleavage signal domain caused a fusion-delayed phenotype (Navas-Martin, Hingley, and Weiss, 2005). The R654D/E1035D suppression was confirmed because the recombinant mutant spike Q159L/R654D/H716D/E1035D has the same virulence as that of the revertant Q159L/R654D/H716D/E1035D (Navas-Martin, Hingley, and Weiss, 2005). The virus containing the single mutant Q159L can replicate efficiently in the brain, but not in the liver (Navas-Martin, Hingley, and Weiss, 2005). However, the mutant Q159L/R654D/H716D/E1035D has no such phenotype in the brain (Navas-Martin, Hingley, and Weiss, 2005), indicating that the revertant may have a different spike protein property. This suggests that these second-site suppressors may have an influence on the functions of the spike protein associated with the pathogenesis of murine coronavirus strain A59.

The yellow fever 17D virus (YF17D) variant, which carries the single substitution D360G in the envelope protein (E protein), causes a persistent infection in mouse neuroblastoma cells with defective cell penetration and small plaque size (Vlaycheva et al., 2004). After YF17D RNA transcripts were transfected in Vero cells, intermediate- and large- plaque revertants were

isolated and determined to be pseudo-revertants, which contain two second-site suppressors A261V and K303N together in the E protein (Vlaycheva et al., 2004). Pseudo-revertant D360G/A361V/K303N has the plaque size, cell growth, and cell penetration similar to the wild type virus, implying that pseudo-revertant D360G/A361V/K303N restores the defect of the mutation D360G (Vlaycheva et al., 2004).

The mutant human immunodeficiency type 1 (HIV-1), which contains two single substitutions R10A/K11A in the nucleocapsid, is defective in viral replication (Cimarelli et al., 2000). The double mutant R10A/K11A also causes a reduction in virion assembly, the incorporation of the viral genomic RNA into the virion, viral DNA synthesis, and viral infectivity (Cimarelli et al., 2000). After a serial passage of the mutant virus in tissue culture, the second-site suppressor E21K in the nucleocapsid is sufficient to restore the above impaired phenotypes to almost the wild type levels (Cimarelli et al., 2000). The second-site suppressor E21 lies on the surface of the nucleocapsid the same as the residues R10 and K11 (Cimarelli et al., 2000), suggesting that E21K may rescue the local charge on the mutant nucleocapsid surface, which was destroyed by the double mutant R10A/K11A.

### **The second-site suppressors isolated from Rous Sarcoma Virus capsid protein**

The second-site suppressors A38V and P65Q in the N-terminal domain (NTD) of the Rous Sarcoma Virus (RSV) capsid protein (CA) were found to compensate for the lethal mutations, L171V and R170Q, respectively. Both single mutations L171V and R170Q are located in a conserved major homology region (MHR) of the RSV CA protein C-terminal domain (CTD) (Bowzard, Wills, and Craven, 2001; Lokhandwala et al., 2008; Purdy et al., 2008). This suggests a critical physical contact between NTD and the MHR region in the CA protein and may be involved in the NTD-CTD interaction. Furthermore, the suppressor I190V can compensate for the single substitution F167Y, which was constructed to destroy the conserved hydrophobic residues located within the MHR hydrophobic core of the CTD (Purdy et al., 2008). Another second-site suppressor S241L was found to rescue the MHR mutant F167Y in the RSV (Bowzard, Wills, and Craven, 2001; Lokhandwala et al., 2008; Purdy et al., 2008). Interestingly, the suppressor S241L is located in the spacer peptide in the CA protein (CA-SP) which does not present in the mature capsid (Bowzard, Wills, and Craven, 2001), indicating that the MHR of CA protein plays a role in conformational changes in RSV, *e.g.* proteolysis, and the single substitution S241L can restore the lethality of the CA caused by the MHR mutation F167Y (Bowzard, Wills, and Craven, 2001). Nevertheless, four suppressors, R185, I190, C192, and F193, which rescue the single substitutions in the CTD were isolated in a special  $\alpha$ 2-helix region, predicting that the above helix may serve a critical role in capsid assembly as a site of the CTD-CTD dimerization in RSV (Lokhandwala et al., 2008). All of the intragenic second-site suppressions were confirmed by assaying the double mutants L171V/A38V, R170Q/P65Q, F167Y/I190V, F167Y/S241L, D155Y/R185, K5R/C192R, and E162Q/F193. Although not all the suppressors can provide the full reversion, these data suggest that the MHR region and the  $\alpha$ 2-helix region in the CTD

possibly function in the conformational changes and dimerization during the RSV capsid maturation.

### **The intergenic suppressors isolated from the *E. coli* bacteriophages Lambda, ΦX174, and filamentous phage f1**

Intergenic suppressors are the other commonly isolated suppressors, since usually proteins need to interact with others to generate the network interactions. Mutations in the bacteriophage lambda gene *FI* are leaky. Lambda produces very few phage particles when there is no functional gpFI. It was believed that gpFI stimulates one of the intermediates for capsid assembly, called complex II, containing a procapsid, a terminase, and DNA (Murialdo and Tzamtzis, 1997). The intergenic second-site suppressors of the mutant gpFI called “fin” (for FI independence) have been isolated individually from the coat protein gene, *E*, and the terminase genes, *NuI* and *A* (Murialdo and Tzamtzis, 1997; Murialdo et al., 1997). Some of the above revertants were found in a region of the gpNuI/A terminase complex encoded near the gene *NuI/A* boundary (Murialdo et al., 1997). The other group of fin suppressors was mapped to a gpE region from the position 158 to 183, called the EFi domain (Murialdo and Tzamtzis, 1997). The above fin suppressors were isolated under the gpFI dysfunction condition, indicating that the catalytic protein gpFI can directly interact with the terminase proteins gpNuI/A and the coat protein gpE to facilitate the above complex II formation.

The second-site suppressors were identified in the coat protein gpF of the prohead accessory protein gpB in the *E. coli* bacteriophage ΦX174 (Fane and Hayashi, 1991). Since the original substitutions in the gpB cause the cold-sensitive phenotype, the suppressor substitutions in the gpFs may facilitate assembly of stable intermediate precursor capsids which could be accumulated at lower temperatures (Fane and Hayashi, 1991).

Another example is in the filamentous phage f1. The suppressor L298F in the protein pI was isolated from the filamentous phage f1 ts mutant pIV (=the minor coat protein) M5S,



(Russel, 1993). This suggests a direct interaction between the proteins pI and pIV. The putative interaction between the proteins pI and pIV may require the carboxyl-terminal domain of pI and the amino-terminal domain of pIV where the mutations locate within (Russel, 1993). Interestingly, the intergenic suppressor in the phage f1 coat protein pVIII was isolated under the background of the phage IKe amI amIV, which contains amber mutations in gene I and gene IV in the phage IKe, with the support of the phage f1 proteins pI and pVI (Russel, 1993), implying the possible interactions of the three proteins pVIII, pI, and pIV.

In the above cases, mutations were isolated from the structural genes either intragenic or intergenic suppressors in the bacteriophages and viruses which have deleterious or defective phenotypes under certain conditions. These mutational studies provided useful *in vivo* information. The functional restoration of the mutated proteins offers convincing evidence that is directly based on the phenotype reversion. By isolating suppressors in the revertants, the protein interactions can be identified and the functions of the target residues (or the target proteins) can be determined. Furthermore, it is not always easy to interpret the results of *in vitro* biochemical experiments and apply them to the protein functions *in vivo*. To overcome the limitations of the biochemical and structural study and to provide a genetic approach *in vivo*, screening suppressors from the mutant HK97 gp5 is a potential way to identify a specific residue (or a particular region) that is critical for one stage of HK97 capsid assembly and maturation. Theoretically, suppressors can be isolated that block at any step during capsid assembly and maturation. Besides, isolating suppressors of the lethal mutations in the gp5 takes advantage of the powerful selection for the restoration of the phage viability, which the genetic changes are sufficient to recover the deleterious mutants.

## 2.0 MATERIALS & METHODS

### 2.1 MATERIALS

#### **Bacterial strains and lysogens**

The *E. coli* bacterial strain DH10B was used to carry the capsid protein expression plasmids. The strain BL21(DE3) was used for protein expression. All of the HK97 lysogens were derived from the strain BW25113 (Table 1). The *E. coli* strain, Ymel, was used for the indicator bacteria. All recombineering strains were constructed by Danju Tso except for the parent strain BW25113 (from Dr. Don Court). All HK97 lysogens were constructed by Danju Tso.

#### **Plasmids**

The pRED plasmid, pKD46, was used to promote homologous recombination in HK97 lysogen. The temperature sensitive plasmid pKD46 carries all three lambda *red* genes: *exo*, *bet*, *gam*, controlled by a P<sub>ara</sub> promoter (Guzman et al., 1995). The permissive temperature for the strain carrying the pKD46 is 30-32°C.

All plasmids which encode the mutant gene 5 were constructed starting from either plasmid pV0 (Figure 8) or plasmid pVB. Plasmid pV0 carries a truncated gene 3 (portal protein gp3) and intact gene 4 (protease gp4) and gene 5 (major capsid protein gp5), all driven from a phage T7 promoter. Plasmid pVB is identical to pV0 except that it carries an engineered

frameshift in gene 4, which inactivates the protease {Duda et al., 1995}. Expression of plasmid pVB leads to efficient production of HK97 Prohead I (uncleaved gp5) and expression of plasmid pV0 leads to efficient production of HK97 Prohead II (cleaved gp5). All mutant plasmids were constructed by Danju Tso except for pV0-Δloop and pV0-D231A (from Dr. Li), and pV0-K178, pV0-K178 (from Dr. Baros).

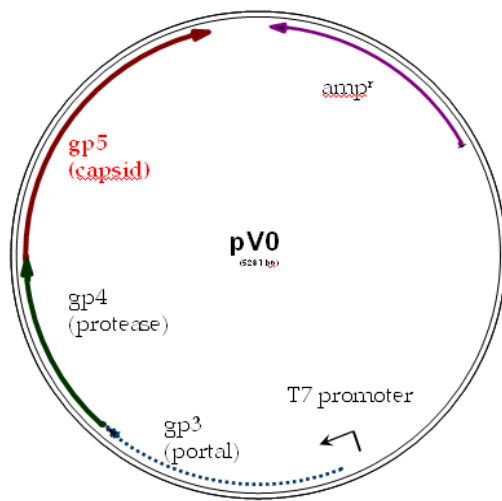
**Table 1: The recombineering bacterial strains and lysogens**

Name	Genotype
BW25113 <sup>@</sup>	W3110-derived recombineering strain: Δ(araD-araB)567, ΔlacZ <sub>4787</sub> (::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, rrnB-3, hsdR514
DT001	BW25113, pKD46, Δ <i>galK</i>
DT002	BW25113, pKD46, Δ <i>galK</i> , HK97
DT003(Kan)	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5<galK, KanR The <i>Kan</i> gene was inserted between HK97 gene <i>tfa</i> and <i>sib</i>
DT004	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5<galK #
DT005	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5<galK *
DT-HK97	BW25113, Δ <i>galK</i> , HK97
DT-SM1	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-A164A
DT-SM2	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-P279P, R280R
DT-V163D	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-V163D
DT-D231L	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-D231L
DT-L224Q	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-L224Q
DT-E165R	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-E165R
DT-E165L	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-E165L
DT-E165K	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-E165K
DT-R365D	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-R365D
DT-N278R	BW25113, pKD46, Δ <i>galK</i> , HK97: gp5-N278R

<sup>@</sup> The strain was supplied by Dr. Don Court.

\* The deletion region in HK97: from the gp4 residue G224 to the gp5 residue Q264

# The deletion region in gp5: from gp5 residue D256 to P373



**Figure 8: The genetic map of the plasmid pV0**

## **Media**

### Luria Broth

10g Bacto tryptone, 5g Bacto yeast extract, and 5g NaCl in 1L ddH<sub>2</sub>O for 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 0.5% (w/v) NaCl. Antibiotics, e.g. 50 µg/ml ampicillin, were added after autoclaving. LB plate was used the same as the above plus 15g Bacto agr for 1.5% (w/v) Bacto agar.

### Soft agar

1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 0.5% (w/v) NaCl, and 0.7% (w/v) Bacto agar.

### Tryptone plate

1% (w/v) Bacto tryptone, 0.5% (w/v) NaCl, and 1.1% (w/v) Bacto agar.

dMAM M9 glucose medium (per 1L): Difco™ Methionine Assay Medium, 1x M9 salts, 0.4% glucose, 1mM MgSO<sub>4</sub>, 10µM CaCl<sub>2</sub>. Antibiotics, e.g. 50 µg/ml ampicillin, were added after autoclaving.

### MacConkey agarose plate (1% galactose) (per 1L)

40g MacConkey agar was added into 950ml ddH<sub>2</sub>O and autoclaved without galactose. Galactose was added after the agar was cooled to approximately 60°C. The galactose was prepared in 20% stock and was sterilized by filtration

#### M63 (0.2% galactose) agarose plate (per 1L)

1x M63 salts, 1 mg/L D-Biotin, 45 mg/L L-Leucine, 0.5 µg/ml FeSO<sub>4</sub>, 0.2 % D-Galactose. All are shown in final concentration and all of the above components were sterilized by filtration. 15g agarose was added into ddH<sub>2</sub>O for autoclaving. All of the sterilized components were added after the agar was cooled to 60°C. M63 salts can be replaced by M9 salts.

#### M63 (0.2% deoxy-galactose) agar (per 1L)

1x M63 salts, 1 mg/L D-Biotin, 45 mg/L L-Leucine, 0.001 % Thiamine(Vit.B1), 0.2 % Glycerol, 2g deoxy-galactose. All of the above concentrations are shown in final concentration and all of the above components were sterilized by filtration. 15g agarose was added into ddH<sub>2</sub>O for autoclaving. All of the sterilized components were added after the agar was cooled to 60°C. The M63 salts can be replaced by the M9 salts.

#### **Buffers & Reagents**

10x concentrated M63 salts (per 100ml): 3g KH<sub>2</sub>PO<sub>4</sub>, 7g K<sub>2</sub>HPO<sub>4</sub>, 2g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>

10x concentrated M9 salts (per 100ml): 13.3g Na<sub>2</sub>HPO<sub>4</sub>· 7H<sub>2</sub>O, 3g KH<sub>2</sub>PO<sub>4</sub>, 1g NH<sub>4</sub>Cl, and 0.5g NaCl

TMG phage buffer: 0.01% gelatin, 10mM Tris-Cl (pH 7.5), 10mM MgCl

UV light induction buffer: 10mM Tris-Cl (pH 7.5), 5mM MgSO<sub>4</sub>

High salt/low pH (expansion) buffer: 50mM NaOAc, pH4.1-4.2, 400mM KCl

DMF (expansion) buffer: 40%(v/v) DMF TAMg buffer

4x sample buffer for SDS PAGE: 0.25M Tris-Cl (pH 6.8), 40% glycerol, 20% beta-mercaptoethanol, and 8% SDS

4x sample buffer for native polyacrylamide gel: 0.25M Tris-Cl (pH 6.8), 40% glycerol, and 20% beta-mercaptoethanol

10x running buffer for native polyacrylamide gel: 0.23M Tris base and 2.5M glycine

10x running buffer for SDS PAGE: 0.23M Tris base, 2.5M glycine, and 1% SDS

Coomessi staining solution: 2.5g Coomessie Brilliant Blue R-250, 500ml 10% acetic acid, 50% MeOH

4x lower buffer for SDS PAGE: 1.5M Tris-Cl, pH 8.8, 0.4% SDS

4x upper buffer for SDS PAGE: 0.5M Tris-Cl, pH 6.8, 0.4% SDS

4x lower buffer for native polyacrylamide gel: 3M Tris-Cl, pH 8.8

4x upper buffer for native polyacrylamide gel: 0.5M Tris-Cl, pH 6.8

TKG50: 20mM Tris-Cl (pH 7.5), 50mM potassium glutamate

50x concentrated TAE running buffer: 1M Tris-acetate, 50mM EDTA pH 8.0

50x concentrated TAMg running buffer: 242g Tris base, 57.1ml acetic acid, pH 8.1, 12.3g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$

KG: 1x M9 salts, 5mM  $\text{MgSO}_4$ , 1.5% casamino acids and 0.2% glucose in ddH<sub>2</sub>O

Lysis Buffer: 50mM Tris-Cl pH 8.0, 5mM EDTA

DL buffer: 20mM Tris-Cl pH 7.5, 40mM NaCl



## 2.2 METHODS

### Construction of HK97 lysogen

The *E. coli* strain BW25113 (pKD46) was grown in LB (amp100µg/L) medium at 30°C until OD<sub>550</sub>~0.3 ( $2.0 \times 10^8$  cfu/ml, exponential phase). BW25113 was infected by the bacteriophage HK97, M.O.I=20, and incubated at 30°C for at least four hours. The above culture was streaked out on LB (amp100µg/L) plate to obtain single colonies.

### Immunity assay

An immunity assay was used to test for the presence of a prophage HK97. Bacteriophage lambda has the same immunity as HK97 as a lambda clear plaque variant, λb2cI, which can be used to assess the presence of a prophage HK97 by sensing the cell's immunity. The isolated lysogen candidate was cross-streaked with λb2cI on an LB plate and incubated at 37°C, overnight. The phage has no apparent effect on the growth of a HK97 lysogen.

### Preparation of bacteriophage plate stocks

HK97 was plated on a lawn of *E. coli* Ymel (see Chapter 2, Material). A single plaque was picked by removing an agar containing a single plaque and mixed with 1ml 1x TMG buffer. The sample was incubated at 37°C, >1hr, to enhance the phage diffusion from the agarose. 0.1 ml of the resuspended phage was mixed with 0.2 ml of an overnight Ymel culture and the mixed sample was incubated at 37°C, >15', for phage absorption. The above infected cells were mixed

with 2 ml soft agar and 2 ml KG medium for plating on a fresh Tryptone plate. The plate was incubated for 6-7 hrs at 37°C until confluent lysis. The top soft agar was collected into a 50ml plastic tube. 2-3 ml 1x TMG buffer and 0.2 ml chloroform were added for phage release. The tube was vortexed vigorously for 20 sec. and held at room temperature (=R.T.) for 20 min. The phage supernatant was separated from the soft agar by centrifugation at JA18 rotor 8K, 10 min and kept at 4°C.

### **Colony PCR and phage PCR**

A single overnight grown colony was suspended in 20µl distilled water and 1µl of the suspension was added to the reaction as the template. For plaque PCR, 1µl of the phage stock ( $\geq 10^8$  pfu/ml) was added into the reaction as the template. The primers binding to gene 4 (forward), *e.g.* DRHpl3, and gene 6 (reverse), *e.g.* gp6-r1 and gp6-r2, were used for amplifying the HK97 gene 5. The PCR product was purified by gel extraction and digested by restriction enzyme to confirm the mutation in gene 5. The entire gene 5 sequence was confirmed by sequence analysis.

### **Site-directed mutagenesis**

The pairs of the mutagenic primers (forward and reverse) (Table 2) contain the designed mutations to substitute a single residue in HK97 gp5. To assist a quick screening of mutant plasmids, a restriction site at (or nearby) the target residue was created or mutated without influencing the designed mutation. The nucleotide sequence of the mutant gene 5 was confirmed by sequence analysis.

**Table 2: The primer pairs designed for constructing the single mutations in HK97 gp5**

Primer name	Direction	Sequence
V163Df	forward	GCCGACGTC <u>GAT</u> GCAGAGAA
V163Dr	reverse	TTCTCTGCATCGACGTCGGC
D231Lf	forward	GGTACCGGTCTCAACCTGGA
D231Lr	reverse	TCCAGGTTGAGACCGGTACC
E165Rf	forward	GTGGTGGCCCGCAAAGCACTGAAG
E165Rr	reverse	TGCTTTGCGGGCCACCACGTCGGC
E165Kf	forward	CGCCGACGTGGTGGCCAAGAAAGCACTGAAGCCAG
E165Kr	reverse	GCTTTCTTGGCCACCACGTCGGCGTTATTGGTAAAC
E165Lf	forward	CGTGGTGGCACTTAAGGCAC
E165Lr	reverse	GTGCCTTAAGTGCCACCACG
L224Qf	forward	GGAAGAGGGCCAGCTTCAAACGGCG
L224Qr	reverse	CGCCGTTTTGAAGCTGGCCCTCTTCC
N278Rf	forward	TCGTCCTGCGGCCTCGCGACTG
N278Rr	reverse	CAGTCGCGAGGCCCGCAGGACGA
R365Df	forward	GCGAAGAAGATCTGGCGCTG
R365Dr	reverse	CAGCGCCAGATCTTCTTCGC
SM1f	forward	GTGGTGGCCGAGAAAGCACTG
SM1r	reverse	TGCTTTCTCGGCCACCACGTCG
SM2f	forward	CCTGAACCCCCGGGACTGGC
SM2r	reverse	GCCAGTCCCGGGGGTTCAGG

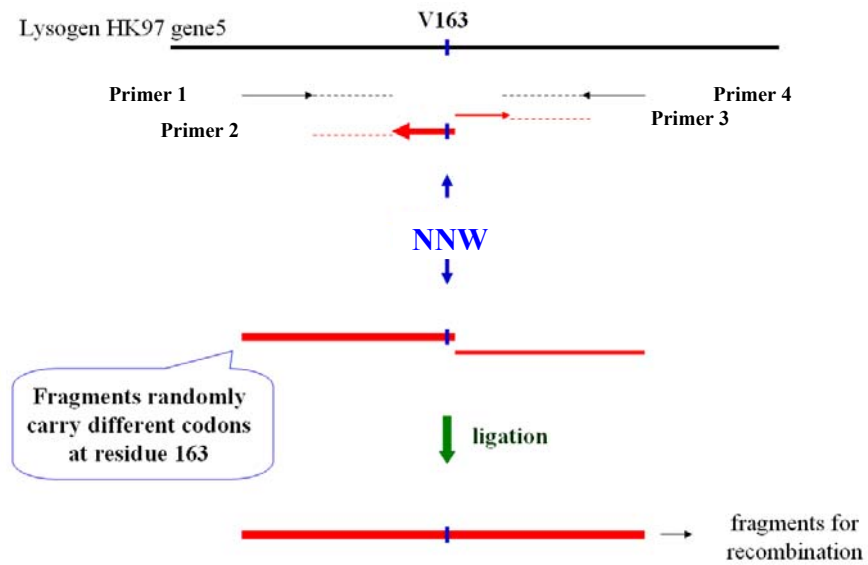
(The substituted nucleotides are underlined)

### **Generation of a selection tag in the mutant lysogen HK97**

In order to establish a quick way to screen the revertants and to rule out the contamination by wild type plasmid or bacteriophage HK97, a silent mutation was generated next to the lethal mutation. In all cases, the silent mutation generated a restriction polymorphism that would be reversed in either the case of a simple reversion to wild type or appearance of a contaminant.

### **Generation of dsDNA with randomly mutated gp5**

The pairs of the primers (forward and reverse) contain the designed mutations to generate random single substitutions at the specific residue 163 in gp5 by using the strategy which was described in Adereth et al., 2005. The sequences at residue 163 were synthesized by using the primer, which contains nucleotides NNW (N: any nucleotide, W: A or T nucleotide) to give the mixed amino acids at residue 163. The amplification of the DNA fragment was manipulated into two steps: First, amplifying two separate DNA fragments which contain two half-length gp5 (Figure 9). Second, the above two fragments were ligated to generate a full-length gp5. The dsDNA fragment was amplified from the ligated gp5, and the PCR product was purified by gel extraction for subsequent homologous recombination. The sequence of the inner forward primer, gp5f1, is GCAGAGAAAGCACTGAAGCCAG (the primer 3 in Figure 9). The selection tag (see the above description) was generated in gp5f1. The sequence of the inner reverse primer, V163Xr, is WNNCACATCGGCGTTATTGGTAAACACC (the primer 2 in Figure 9). Two of the outer primers are LP1 (forward primer) and gp6-r1 (reverse primer) (the primer 1 and 4 in Figure 9). The outer primers can be substituted with other proper designed primers.



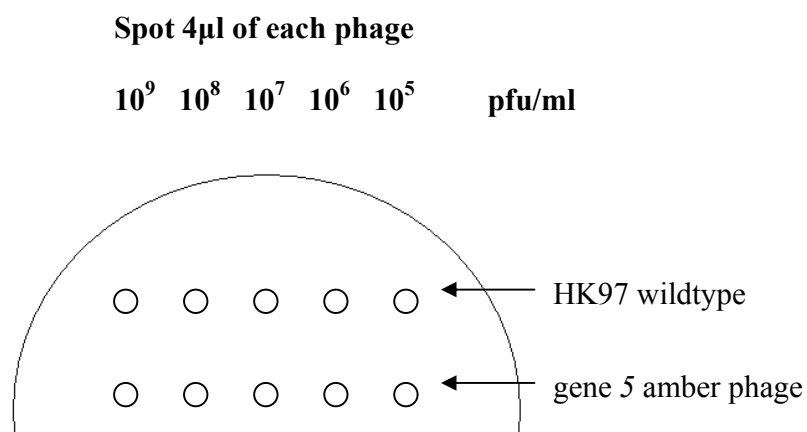
**Figure 9: The strategy to introduce quasi-random amino acids at gp5 residue 163**

The DNA fragment containing quasi-random amino acids at gp5 residue 163 was amplified by LP1 (labeled primer 1) and V163Xr (labeled primer 2). The rest gene 5 sequence was amplified by gp5f1 (labeled primer 3) and gp6-r1 (labeled primer 4). The above two DNA fragments were ligated to generate a full-length gp5 and the linear dsDNA fragment (gp5) was amplified from the ligated full-length gp5 for subsequent homologous recombination.

### Complementation spot test

The plasmid encoding the mutant gp5 to be tested for functionality was introduced into the strain BL21(DE3), pLysS. A serial dilution of the phage carrying an amber mutation in gene 5(=gene 5<sup>-</sup> amber phage) was spotted on the above bacteria lawn which contains the constructed plasmid (Figure 10). Plaques can be produced when the gene 5<sup>-</sup> amber phage can be complemented by the mutant gp5 that is expressed in the strain.

Before recombining a lethal mutation into the HK97 prophage for constructing a mutant lysogen, the mutations were examined for their gp5's biological activity. Mutant gp5's, which show less than 10<sup>3</sup> fold in the complementation test, were categorized as lethal mutations. Others were not introduced into the lysogens for isolating suppressors since they may be leaky. For instance, the mutant gp5 R280I, which carries a single substitution R280I in gp5, has about 100 fold less than the wild type (Appendix A). So R280I is not counted for lethal mutation and the mutant lysogen DT-R280I is leaky.



**Figure 10: The phage complementation assay**

### **Recombination in lysogens**

The HK97 lysogen was incubated in LB (amp100 $\mu$ g/L) medium with the addition of 10mM L-arabinose until OD550  $\geq$ 0.4. This allows bacteria to accumulate the lambda Red proteins for homologous recombination. Lysogens were harvested, chilled on ice for 10 min, and samples were centrifuged at 6700rpm for 10 min, at 4°C, followed by washing three times with pre-chilled distilled water. These recombineering-competent cells were suspended in pre-chilled ddH<sub>2</sub>O. A particular mutant gene 5 was amplified from one mutant plasmid pV0 and introduced into the recombineering lysogen via electroporation. The electroporated cells are resuspended in LB (amp100 $\mu$ g/L) medium with the addition of 10mM L-arabinose. The step allows cells to recover from the electroporation and keep expressing the functional Red proteins. Cells are cultured at 30°C for ~ 3 hours followed by washing three times with 1x minimal salts (M63 or M9) to remove the LB medium. All cells are kept in a small amount of 1x minimal salts and plated on the special selection plates.

There are two controls for the homologous recombination experiment: 1) the lysogen HK97 without L-arabinose induction but with the transformation of the mutant gene 5. 2) The lysogen HK97 with L-arabinose induction but without the transformed mutant gene 5. There should be no colony appearing on these control plates after incubation. Recombination efficiency was around 10<sup>-6</sup> but dropped ~10 fold when the frozen recombineering lysogens were used.

**Table 3: The primer pairs designed to amplify the *galK* gene for subsequent homologous recombination**

Primer name	Direction	Sequence
HRgalKf	forward	<u>AGCAGAAATCGAAAGCACAGGCCAGGTTTCCAAACAGTT</u> <u>GCAGTCCGACCCCTGTTGACAATTAATCATCGGCA</u>
HRgalKr	reverse	<u>GGGTTTCAGGACGATAACCGGAAGCGCTGAACTCAGACTCG</u> <u>GTCACCTGTCAGCACTGTCCTGCTCCTT</u>
HRgalKf-1	forward	<u>GCACAAATAAACGCTCTGCTTCAGAGCATCAAATCTTTCC</u> <u>CTTCTAACTTAGGCCTGTTGACAATTAATCATCGGCA</u>
HRgalKr-1	reverse	<u>CCGCCCCCTCCATCATGAGCCAGAAGAGAAGGTGCCCTT</u> <u>GATGATTGCTGTCAGCACTGTCCTGCTCCTT</u>
HRgalKf'	forward	<u>GTGGCAACCGCCTATGACACCTCGCTGAATGCCACCGGC</u> <u>GACACCCGCGCCCTGTTGACAATTAATCATCGGCA</u>

(The 5' end underlined sequence is the HK97 gp5 nucleotide sequences served as the homologous region. The 3' end sequence is the designed primers for amplifying the *galK* gene)



### **Generation of double mutant plasmids**

HK97 double mutant gp5's were amplified from the HK97 lysogens D231L/K178V, D231L/K178I, and D231L/K178N. The amplified fragments were digested with *EcoRV* and *BstEII* and the plasmid pV0 was digested with the same two enzymes. Mutant plasmids pV0-D231L/K178V, pV0-D231L/K178I, and pV0-D231L/K178N, were generated by ligating the plasmid pV0 with three different double mutant DNA. Three mutant plasmids pVB-D231L/K178V, pVB-D231L/K178I, and pVB-D231L/K178N (lacking the protease gp4) were generated based on the same strategy.

### **Ultraviolet irradiation**

The mutant lysogen culture grown until the early exponential phase was harvested and spun down for UV light induction (the intensity is 6 unit if read by the long wave meter (J-221)). A standard burst size was determined the control strain, DT-HK97. DT-HK97 was treated with variable doses of ultraviolet irradiation, *i.e.* 0 sec, 1 sec, 2 sec, 3 sec, 4 sec, 5 sec, 10 sec, 15 sec, 20 sec, 25 sec, 30 sec, 40 sec, and 50 sec, to test the best condition for inducing the control lysogens. After exposure, lysogens were harvested and incubated in LB medium at 37°C for phage production. The supernatant was collected and the phage titer was determined. The lysogen DT-HK97 achieved the largest burst size when treated for five seconds (data not shown). Similar results were seen for the lysogens DT-SM1 and DT-SM2, which are the lysogens HK97 except for carrying a silent mutation in the gene 5 (data not shown). The HK97 phages were grown at the permissive temperature, 37°C, (and 30°C) after induction. The plaques grown on the indicator bacterial lawn were randomly selected for individual experiments. The presence of

the mutations and the suppressors were verified by restriction polymorphism and sequencing analysis.

### **Electrophoresis (Protein)**

Native agarose gel (0.8%) was used to detect the presence of the assembled proteins, such as capsomers and proheads. Mobility shift occurs on the agarose gel when the size or the surface charge of the capsomers and proheads was changed.

Native polyacrylamide gel (native gel) was used to determine the formation of the pentamers and hexamers. The upper stack: 4% acrylamide (30% acrylamide/0.44% bis-acrylamide) in 0.25M Tris-Cl, pH 6.8. The lower gel: 8% acrylamide (33.5% acrylamide/0.3% bis-acrylamide) in 0.75M Tris-Cl, pH8.8.

Sodium dodecyl sulfate (SDS) polyacrylamide gel (PAGE) was used to detect the gp5 production, and to monitor the protease cleavage on gp5. The sample was collected and precipitated by 10% cold trichloroacetic acid (TCA) for 10min. The TCA sample was washed with acetone for three times to remove the TCA, and the protein pellet was dried up followed by suspension with 1x SDS loading buffer. The sample was boiled for ~3 min to denature the proteins before loading on an SDS PAGE.

All gels were stained in Coomassie Blue and destained in 10% acetic acid to detect the protein presence. For radio-labeled protein expression assays, gels were dried and imaged by radiography.

### **Purification of HK97 Prohead I**

The expression strain BL21(DE3), pLysS, carrying the plasmid pVB, was incubated in one liter LB medium at 37°C, overnight. The HK97 protein gp5 was induced by 0.4mM IPTG at 28°C, overnight. Bacteria were chilled on ice for 10 min, and spun down in the Beckman JA-10 rotor (5K/15min/4°C). Triton X-100 (~0.4%) and 90 ml lysis buffer were added to resuspend the pellet. Samples were heated in the water (22-24°C) for 1 min and chilled on ice for 5 min until they reached ~10°C. The above steps were repeated two more times to improve lysis. MgSO<sub>4</sub> (7.5mM) and DNase I (20µg/ml) were added for digesting DNA, and the samples were incubated at 24°C (R.T.) for 10 min. The pellet was spun down in the JLA-16.25 (10K/15min/4°C) and the protein samples were usually in the supernatant. To precipitate the protein, 0.5M potassium glutamate and 6% PEG were added, and samples were incubated on ice for more than 20 minutes or longer. The PEG precipitation (pcpt) was collected by a spindown in the JLA-16.25 (8K/15min/4°C) and the sample pellet was resuspended in 15 ml TKG buffer. Samples were spun down once in the JLA-16.25 (8K/15min/4°C). The protein pellet was covered with about 10ml of TKG buffer and kept at 4°C, overnight, to resuspend the proheads. Velocity sedimentation was used to purify the above protein in 10-30% glycerol gradient. The protein band was extracted from the gradient, and dialyzed to remove the glycerol. Ion exchange chromatography was performed to further purify the proteins. The fractions, which contain Prohead I, were collected and concentrated by ultracentrifugation using rotor Ti45 (35K/2hrs/4°C). The purified proheads were kept at 4°C for analysis on a native agarose gel and an SDS PAGE.

## **Purification of HK97 Prohead II**

The expression strain BL21(DE3), pLysS, which carries the plasmid pV0, was incubated in one liter LB medium at 37°C, overnight. After overnight culture, expression of the HK97 protein gp5 was induced by 0.4mM IPTG at 28°C, overnight. Bacteria were chilled on ice for 10 min, and spun down in the JA-10 (5K/15min/4°C). Triton X-100 (~0.4%) and 90 ml lysis buffer were added to resuspend the pellet. Samples were warmed in water (22-24°C) for 1 min and chilled on ice for 5 min until ~10°C. The above steps were repeated for two more times to release the proteins from bacteria. After MgSO<sub>4</sub> (7.5mM) and DNase I (20µg/ml) were added, the samples were incubated at 24°C (R.T.) for 10 min for digesting DNA. The samples were spun down in the JLA-16.25 (10K/15min/4°C) and the proteins were usually in the supernatant. To precipitate the protein, 0.5M NaCl and 6% PEG were added. The samples were incubated on ice for more than 20 min. The PEG precipitation (pcpt) was collected by a spindown in the JLA-16.25 (8K/15min/4°C). The protein pellet was resuspended in 15 ml DL buffer and the protein samples were spun down in the JLA-16.25 (8K/15min/4°C). The sample pellet was covered by little amounts of DL buffer (~10ml) and kept at 4°C, overnight, to resuspend the proheads. Velocity sedimentation was used to purify the above protein in 10-30% glycerol gradient. The protein band was extracted from the gradient, and dialyzed to remove the glycerol. Ion exchange chromatography was performed to further purify the proteins. The fractions, which contain Prohead II, were collected and concentrated by ultracentrifugation in the rotor Ti45 (35K/2hrs/4°C). The purified proheads were kept at 4°C for analysis on a native agarose gel and an SDS PAGE.

### **Prohead II expansion assay**

Working stock of 20mg/ml Prohead II was diluted 1:10 into the expansion reaction buffer (2mg/ml in final concentration) at 25-26°C. Samples were neutralized to stop the reaction by diluting the samples into DL buffer. All samples were analyzed on an agarose gel.

### **Prohead II crosslinking assay**

1) SDS sample buffer induction: working stock 15mg/ml Prohead II was diluted 1:4 into the SDS sample buffer and incubated at R.T. The samples were directly analyzed on an SDS PAGE.

2) High salt/low pH buffer induction: working stock 20mg/ml Prohead II was diluted 1:10 into the high salt/low pH buffer and incubate for one hour. Samples were neutralized by diluting into the DL buffer to induce crosslinking. All samples were analyzed on an SDS PAGE.

3) DMF buffer induction: working stock 20mg/ml Prohead II was diluted 1:10 into the DMF buffer and incubated for one hour. Samples were then neutralized by diluting into the TAMg running buffer to induce crosslinking. All samples were analyzed on an SDS PAGE.

### **Radio-labeled protein expression**

A 50µl turbid culture sample of the expressing strain BL21(DE3), which carries the target plasmid, was diluted in 5 ml LB medium and incubated at 30°C overnight for inducing protein. The above overnight culture was kept incubation at 37°C until OD<sub>550</sub>~0.4. The culture was induced by 0.4mM IPTG for 30 min. After 30 min incubation, 200µg/ml Rifampicin was added to inhibit the host protein synthesis for 15 min. One ml sample was removed into another test tube, and 1µl <sup>35</sup>S-Met (10 mCi/ml) was added to radio-label the gp5. 100µl samples were collected into 500µl 10% TCA at time points 0 min, 5 min, and 45 min after <sup>35</sup>S-Met was added

for analysis on an SDS PAGE. All TCA samples were precipitated at -20°C for at least 10 min. All of the remaining samples were harvested and chilled on ice for 10 min. The spin-down pellet was resuspended in the 500µl lysis buffer. 0.1% Triton-100(TX-100) was added to release the protein from bacteria and samples were warmed at 24°C (R.T.) for 5 min. Tubes were then transferred back on ice and 7.5mM MgSO<sub>4</sub> and 20µg/ml DNase I were added to digest the released DNA. Samples were held at 25-26°C, 10 min, for digestion. The supernatant and the pellet were separated by centrifugating at 12K, 4°C, 10 min followed by analysis on an agarose gel and on a native gel. The sample pellet was resuspended in the 500µl TKG50 and all samples were kept at 4°C or further examined on the native agarose gel and the native PAGE gel.

### **Structural view analysis**

The crystal structures of the HK97 Prohead II, and Head II were downloaded from Protein Data Bank (PDB) (<http://www.rcsb.org>) and analyzed by the Swiss-PDB viewer, DeepView v4.0 (<http://spdbv.vital-it.ch>). The PDB accession number of HK97 Head II (3.44Å) is 1OHG {Helgstrand et al., 2003}. The PDB accession number of HK97 mutant W336F, E-loop truncated Prohead II (3.65Å) is 3E8K {Gertsman et al., 2009}.

### **DNA Sequencing**

Sequence results were from Genewiz (<http://www.genewiz.com>).

### **3.0 GP5 RESIDUE SUBSTITUTION IN HK97 LYSOGENS BY RECOMBINEERING**

#### **3.1 INTRODUCTION**

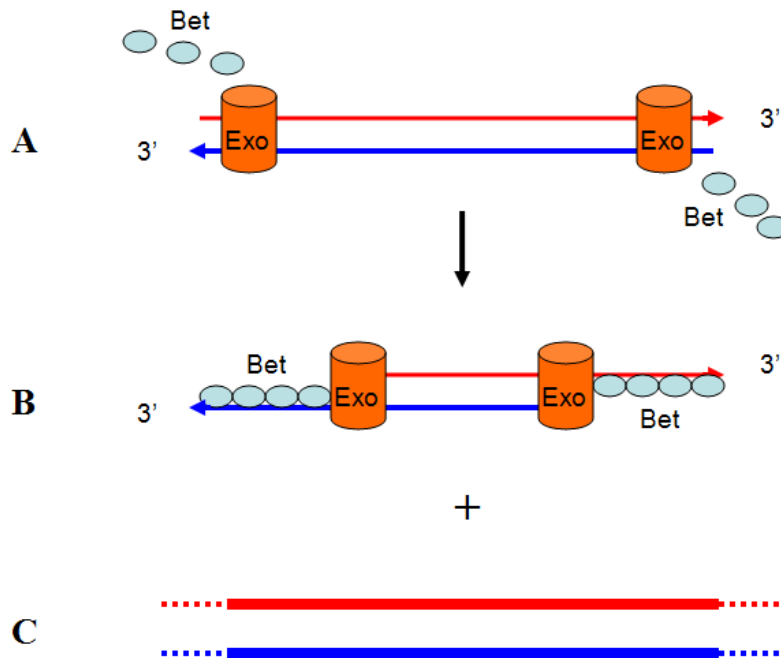
Two closely related bacterial strains can exchange their genes by homologous recombination. Since this type of genetic exchange is sequence specific, homologous recombination has been used to introduce specific DNA fragments into the *E. coli* target gene (Court, Sawitzke, and Thomason, 2002; Datta, Costantino, and Court, 2006; Ellis et al., 2001; Oppenheim et al., 2004; Thomason et al., 2007; Yu et al., 2003). Thus, the genes which are responsible for homologous recombination should be able to introduce DNA fragments into prophages.

In my study, a technique called recombineering was used to integrate any designed mutation into *E. coli* through homologous recombination (Copeland, Jenkins, and Court, 2001; Datta, Costantino, and Court, 2006; Liu, Jenkins, and Copeland, 2003; Oppenheim et al., 2004; Thomason et al., 2007; Warming et al., 2005; Wong et al., 2005; Yu et al., 2003). Recombineering needs very short (~50nt) homologous regions at both ends of the linear DNA (Thomason et al., 2007), providing an opportunity to recombine the lethal mutation into the prophage HK97 gp5 in the lysogens and study different gp5 residues *in vivo*. Lethal mutations cannot be directly introduced into the genome in the bacteriophage capsid. Because of the lysogenic cycle of bacteriophage HK97, a designed mutant gp5 can be introduced into a HK97 lysogen to replace the original residue in gp5.

## Recombineering

Three lambda *red* genes, *exo*, *bet*, and *gam*, are responsible for recombineering-based homologous recombination (Muniyappa and Radding, 1986; Poteete and Fenton, 1993; Poteete, Fenton, and Murphy, 1988; Radding, 1973; Wackernagel and Radding, 1973). Lambda *Exo* protein binds to two free ends of dsDNA fragments and digests a single strand from each end in the 5' to 3' direction (Figure 11). This leaves a linear dsDNA with 3' ssDNA overhangs at both ends (Figure 11) (Copeland, Jenkins, and Court, 2001). The other *red* gene product, *Bet* protein, binds to the ssDNA ends and mediates annealing between them and homologous sequences in the dsDNA target (Figure 11) (Karakousis et al., 1998; Li et al., 1998; Muniyappa and Radding, 1986). Unlike yeast cells, *E. coli* produces a specific enzyme RecBCD, which can destroy exogenous DNA. Since the function of the *gam* protein is to act against the host's RecBCD enzyme, the lambda *gam* protein is required for recombineering (Poteete, Fenton, and Murphy, 1988).

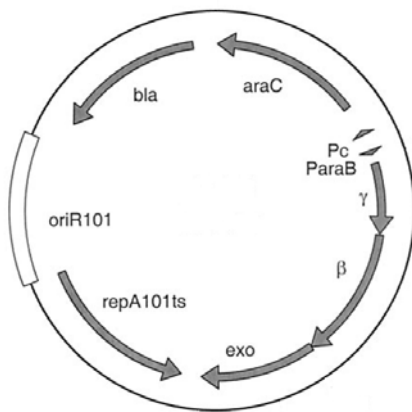




**Figure 11: The lambda Red proteins, Exo and Bet, provide homologous recombination in *E. coli***

Protein Exo recognizes the double-strand linear DNA and binds to its two ends (A) followed by protein Exo digests the DNA from 5' end. The reaction creates a 3' end, single strand DNA (B). Protein Bet recognizes and binds to the linear single strand DNA at two ends and leads it for homologous recombination (B and C).

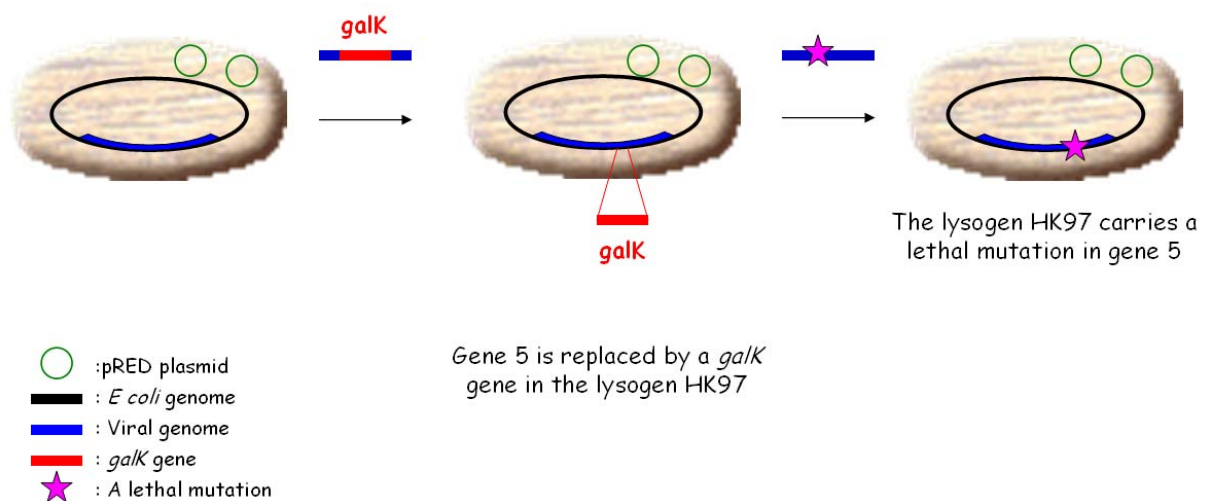
The lambda *red* genes have been adapted for recombineering by two related strategies. In the first, the *red* genes are provided on a defective temperature inducible prophage based on  $\lambda$ cI857 (Yu et al., 2000). In the second, the *red* genes are on the plasmid pRED (Figure 12) (Datsenko and Wanner, 2000). The plasmid pRED is used in my study since the bacteriophage lambda and HK97 have the same immunity and the same insertion site and this makes the first recombineering system incompatible with HK97. In pRED, the three lambda *red* genes are controlled by a specific promoter, called P<sub>araB</sub>. The addition of L-arabinose in the culture can induce the lambda *red* genes (Guzman et al., 1995). Thus, recombination can be easily induced by the addition of L-arabinose to the culture. In most *E. coli* strains, the efficiency of the recombination is one in  $10^5$ - $10^6$  when the plasmid pRED was used (Datsenko and Wanner, 2000).



**Figure 12: The genetic map of the pRED plasmid pKD46 (Datsenko and Wanner, 2000; Guzman et al., 1995)**

### Recombineering-based *galK* selection

To screen for recombinants in the total of  $10^5$ - $10^6$  population is a labor-intensive work. In order to have an efficient way to identify recombinants, a recombineering-based selection system, called *galK* positive/negative selection system, is used in my experiments (Figure 13) (Warming et al., 2005). The *galK* gene product, galactokinase, plays a fundamental role in galactose metabolism by phosphorylating the galactose to galactose-1-P (Debouck et al., 1985) and the *galK* gene is essential when galactose is the only carbon source for *E. coli* (Debouck et al., 1985).



**Figure 13: The recombineering-based *galK* positive/negative selection**

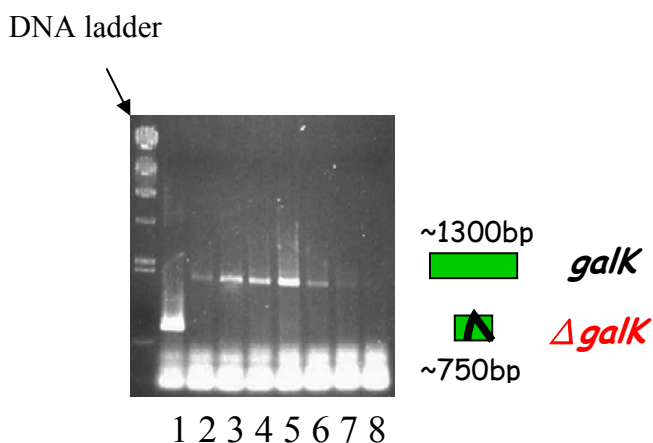
The *galK* positive/negative system is a two-step way to introduce a *galK* gene into the target gene, and, later, the introduced *galK* gene is replaced by the mutated dsDNA fragment (Figure 13). This process is described as follows: First, the *galK* gene with homology at both ends is recombined into the lysogen HK97 to replace the gene 5, and the lysogens are selected on galactose minimal salts plate. Lysogens cannot survive on the above minimal plate if lacking a *galK* gene. This provides a simple step to select the lysogens which carry the *galK* gene. Introduction of a *galK* gene into the gp5 is to generate a selection. Second, the *galK* gene in the lysogen is replaced by a mutated gene 5, which contains the designed mutation (Figure 13). These lysogens are selected on a minimal salts plate which contains 2-deoxy-galactose (DOG). DOG is a synthetic chemical, which can be converted to a toxin to bacteria *E. coli*, yeasts, and some mammalian cell lines (Claverie, de Souza, and Thirion, 1979; Platt, 1984; Thirion, Banville, and Noel, 1976). It has no effect unless the *E. coli* carries galactokinase; if DOG is phosphorylated it becomes toxic (Henderson and Giddens, 1977; Henderson, Giddens, and Jones-Mortimer, 1977). Therefore, lysogens which carry a *galK* gene cannot survive on the DOG plate. By using the DOG plate, the lysogens that carry the lethal mutation in the HK97 gp5 can be selected.

Recombineering has been used in *E. coli* for many years. The target gene in the host genome can be sufficiently recombined with the foreign DNA which contains homologous sequences at the ends. Here I use recombineering to introduce lethal mutations in gp5 into the HK97 lysogen.

## 3.2 RESULTS

### Deleting the *galK* gene in lysogen HK97

In order to establish the *galK* selection in a HK97 lysogen, the genomic *galK* gene in the *E. coli* bacterial genome has to be deleted. The recombineering strain SW102 (Warming et al., 2005) has the genome lacking the *galK* gene. The SW102 genomic fragment containing the mutant *galK* gene was amplified and recombined into the recombineering lysogen BW25113 to replace the wild type *galK* gene. The recombinants were selected for the ability to grow on deoxygalactose (DOG) on a minimal salt plate, analyzed by PCR amplification (Figure 14) and sequence analysis.

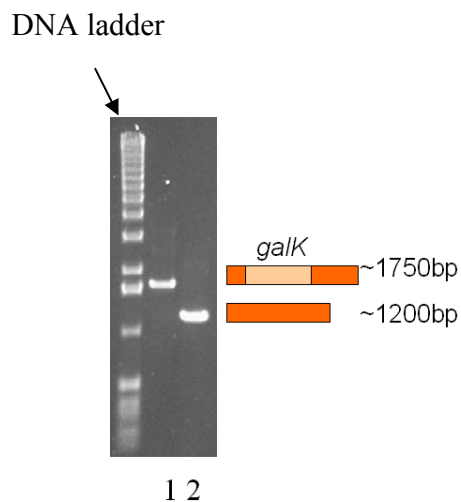


**Figure 14: The amplified products from the colonies grown on DOG plate**

Lanes 1 through 8 are the PCR products which were amplified from the colonies grown on DOG plates. 5  $\mu$ l of the PCR products were analyzed on a 1% agarose gel. Lane 1 shows the fragment without the genomic *galK* gene and lanes 2 to 8 show the fragment with the full length of the genomic *galK* gene.

### Replacing the gene 5 by a *galK* gene in the lysogen HK97

The 200ng linear dsDNA fragment was amplified from the plasmid pGalK, which contains a full-length *galK* gene (Table 3). The primers for this amplification contain HK97 gp5 homologous regions (~80nt) at their 5' ends. The recombinants were selected on a 0.2% galactose minimal salts plate and the *galK* gene in the lysogen was analyzed by PCR amplification (Figure 15) and confirmed by sequencing. All of the selected colonies grown on the above plate are recombinants.



**Figure 15: The amplified product from the lysogens, with or without a *galK* gene in the prophage HK97**

In lane 1, the template is the recombinant grown on a 0.2% galactose plate. In lane 2, HK97 gene 5 was amplified from the lysogen HK97.

### **Recombining the *galK* gene with mutant gene 5 in lysogen HK97**

To generate mutant HK97 lysogens, a mutant gene 5 was amplified to recombine into the HK97 lysogen. The introduced DNA fragment was recombined into prophage HK97 to replace the *galK* gene. The recombinant has a complete gene 5 sequence except for a lethal mutation at one residue. The recombinant with the single substitution can be selected for the ability to grow on DOG on a minimal salt plate. The length of the amplified PCR product from the mutant lysogen is the same as the length of the gene 5, so the recombinant can be easily distinguished from the lysogen which carries a *galK* gene in the HK97 prophage (Figure 15).

To determine the best conditions for selecting recombinants on DOG plates, 100ng, 200ng, and 400ng of the linear dsDNA were introduced into HK97 lysogens individually. The results show that 200ng dsDNA provides the highest efficiency when a ~500nt homologous region was used (Table 4). In another experiment, the same concentrations of DNA (200ng) but with different homologous regions flanking the target region—100nt, 200nt, and 500nt—were introduced into HK97 lysogens individually. The results show that the longer the homologous region, the higher the efficiency (Table 5). Thus, the 200ng dsDNA fragment flanked by 500nt homologous regions at both ends provides the best condition to recombine the mutant gene 5 into the HK97 lysogen.

**Table 4: The recombination efficiency when 500nt length gp5 homologous region and 100ng, 200ng, or 400ng dsDNA were used**

Homologous Length (nt)	DNA Concentration (ng)	Efficiency (%) [Fraction of the colony]
500	100	0-25
500	200	33-85
500	400	9-50

**Table 5: The recombination efficiency when 200ng dsDNA and 110nt, 200nt, or 500nt length of the gp5 homologous region were used**

DNA Concentration (ng)	Homologous Length (nt)	Efficiency (%) [Fraction of the colony]
200	110	7-36
200	200	20*
200	500	33-85

\* Data was not been repeated



The recombination frequency measured in my study is usually one recombinant in  $10^6$  cells (0.0001%) in the absence of selection. It is a complicated work to increase the recombination frequency. However, it is possible to utilize a selection system to increase the “isolation frequency”. In my research, the *galK* which was introduced into the target gene, gene 5, creates a selection to isolate recombinants in a small population grown on the selection plates. The isolation frequency was found at least one recombinant in the five grown cells (20%).

The recombineering-based homologous recombination was provided from the temperature sensitive pRED plasmid, which has temperature sensitive replication, so it is possible to remove the plasmid by growing the lysogens at 42°C. Thus, the mutant lysogens can be used for isolating revertants in the absence of the recombineering plasmid.

## **4.0 THE ROLE OF THE GP5 RESIDUE 163 DURING BACTERIOPHAGE HK97 HEAD PROTEIN CROSSLINKING**

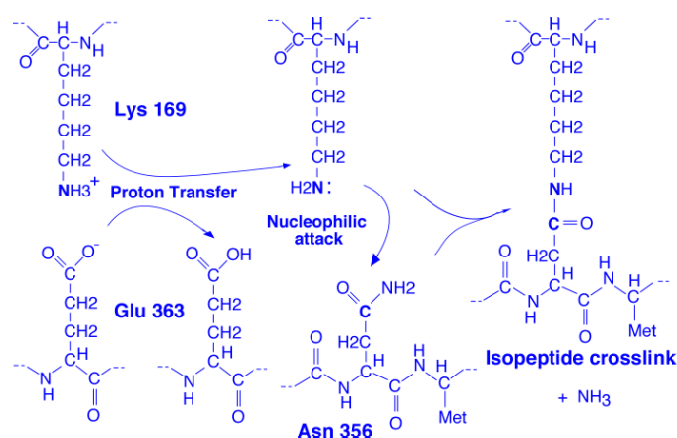
### **4.1 ABSTRACT**

*Escherichia coli* bacteriophage HK97 assembles the mature capsid from 415 copies of the major capsid protein gp5. After assembly in the presence of maturation protease, the precursor capsids called proheads undergo proteolysis and several stages of maturation including expansion and crosslinking to form a mature capsid called Head II. Unlike most other dsDNA bacteriophages, bacteriophage HK97 undergoes head protein crosslinking, forming irreversible isopeptide bonds to secure its capsid. Previous studies show that Lys169 (K169) and Asn356 (N356) form an intersubunit isopeptide bond as the result of the crosslinking, and Glu363 (E363) from a third subunit functions as the catalytic residue to trigger the crosslinking reaction. A chemical mechanism of crosslinking has been proposed in which residue E363 acts as a general base to extract a proton from residue K169, and this reaction can be effective inside a hydrophobic environment. Here I report that the gp5 residue 163 is essential to produce viable HK97 and required for HK97 head protein crosslinking. Genetic data show that residues, Val, Leu, Ile, Thr, or Cys can be tolerated at residue 163 in gp5, suggesting that a non-charged, non-aromatic residue with a certain size of side chain at 163 is enough to produce viable bacteriophages. Biochemical results reveal that mutant gp5 V163D assembles Prohead II efficiently and allows

normal prohead expansion but is specific defect in crosslinking. Our data demonstrate that residue 163 is crucial and essential for HK97 capsid maturation, and specifically plays a role in the head protein crosslink formation.

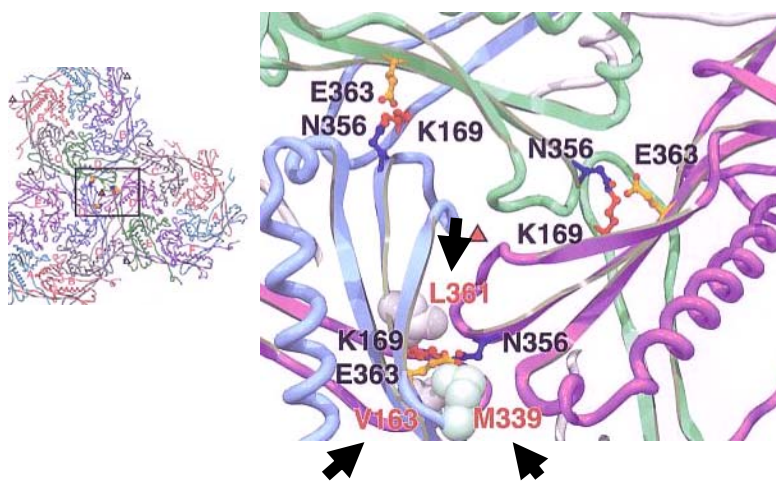
## 4.2 INTRODUCTION

The irreversible covalent crosslink is nearly the last step of bacteriophage HK97 capsid maturation. By forming an isopeptide bond connecting Lys169 (K169) and Asn 356 (N356) from two adjacent gp5's, the crosslinking reaction produces mature HK97 capsid (Duda et al., 1995a). A possible crosslinking reaction mechanism was proposed as follows: Residue Glu363 (E363) extracts a proton from residue K169, producing the neutral amine K169\* (K169\*), then the generated K169\* attacks N356 to form an isopeptide bond between two polypeptide chains in the nearby gp5's (Figure 16) (Dierkes et al., 2009). Such a transfer of a proton is not favored in an aqueous environment, since the pKa values of both Lys and Glu strongly favor their charged forms at the intermediate pH levels (Figure 16) (Dierkes et al., 2009). On the other hand, if the two charged residues were in a hydrophobic environment, the proton transfer would be expected to be much more favorable since it would eliminate both buried charges.



**Figure 16: The proposed chemical mechanism of the HK97 capsid crosslinking**

Three amino acids in the near vicinity of the crosslinking site—V163, L361, and M339—appear from the Head II structure to form a hydrophobic environment surrounding the K169-N356 isopeptide bond (Figure 17) and possibly provide a hydrophobic environment to promote the charge transfer during crosslinking.

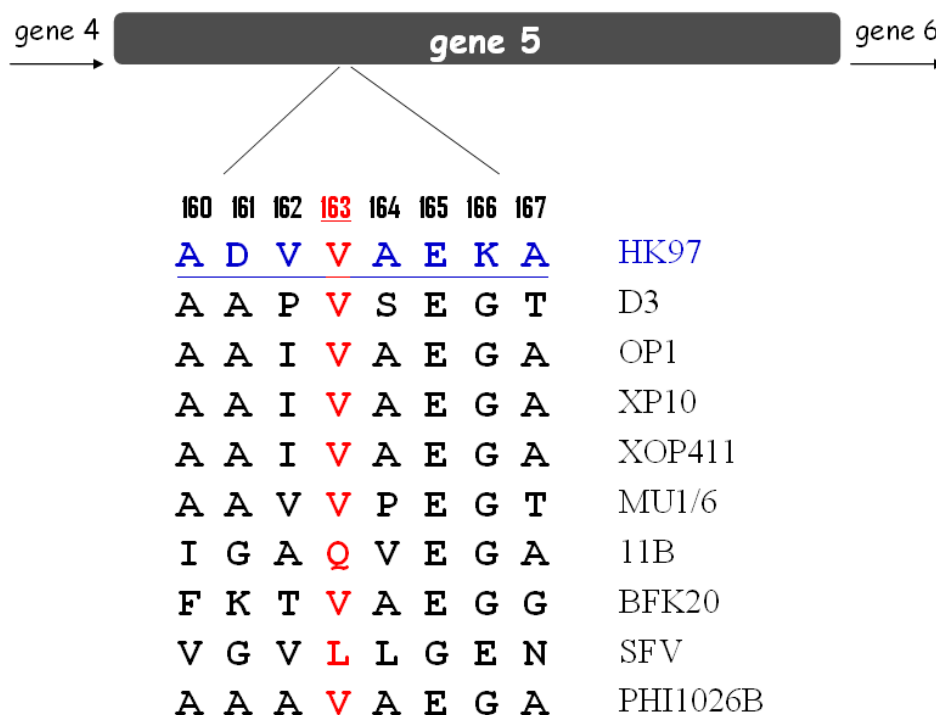


**Figure 17: The HK97 head crosslink site**

The HK97 crosslink sites are located between capsomers as shown on the left figure (Dierkes et al., 2009). The residues, K169, N356, and E363, involved in the crosslinking are shown with their side chains at each crosslink site. A crosslinking reaction results in a K169-N356 isopeptide bond between two adjacent gp5's in two capsomers. The residue E363 located in the third isopeptide bond plays a catalytic role in the crosslinking reaction. Three of the hydrophobic residues, V163, M339, and L361, are found very close to the crosslink site and are indicated by the black arrows on the bottom crosslink site.

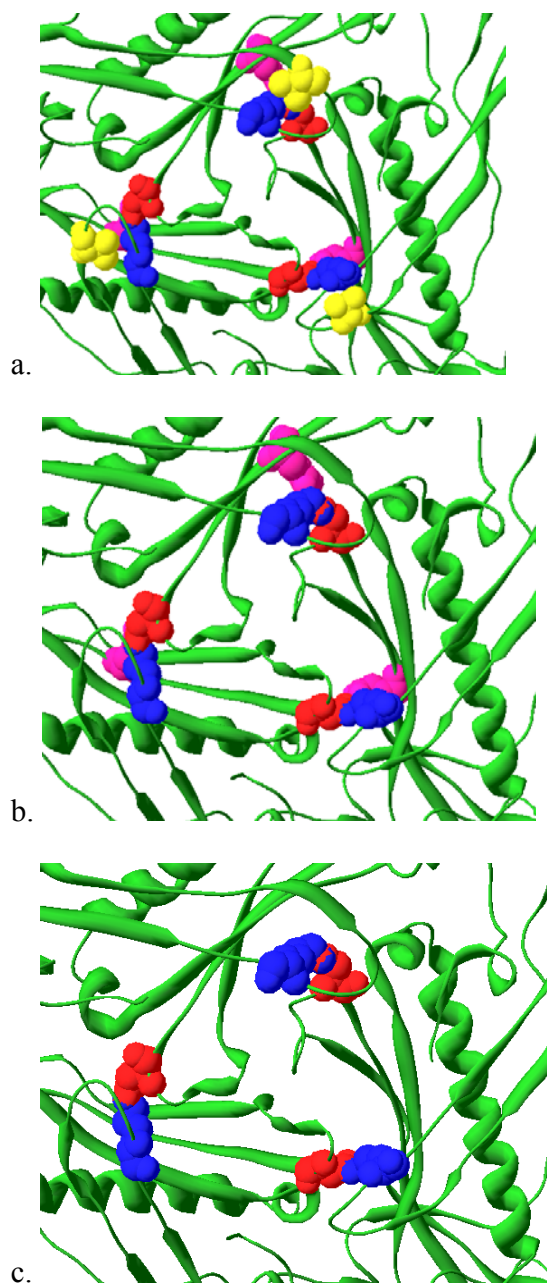
It has been shown that mutant Prohead II K169Y, K169N, N356D, and E363D, E363A, E363H, E363N, and E363Q, cannot crosslink or cause certain defects in the early assembly (Dierkes et al., 2009; Duda et al., 1995a), indicating that they are involved in the head crosslinking reaction. However, the catalytic roles of three residues that flank the active site, V163, M339, and L361, have never been examined.

To dissect the crosslinking mechanism, my studies focus on residue V163. Among these three hydrophobic residues, residue V163 is conserved in many dsDNA bacteriophages whose major capsid proteins have similarities to HK97 gp5 (Figure 18, Appendix B). Unlike residue V163, residue L361 is partially conserved and residue M339 is not conserved in dsDNA bacteriophages (Figure 18, Appendix B). Furthermore, residue V163 is located in the E loop (for Extended loop), a distinct long  $\beta$ -hairpin of gp5 (Figure 4 and 17, and 19). The gp5 E loop undergoes a movement during prohead expansion, from protruding above the capsid shell to lying on the head surface. Therefore, residue V163 in the E loop is translocated to the crosslink site. I hypothesize that residue V163 has a particular role in the head protein crosslinking.



**Figure 18: The partial sequence alignment of the major capsid proteins from the dsDNA phages HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B**

The major capsid proteins of nine dsDNA bacteriophages have similarities to *E. coli* bacteriophage HK97 gp5 (385a.a.). The nine bacteriophages are *Pseudomonas* D3 (395a.a.), *Xanthomonas* phage OP1 (390a.a.), *Xanthomonas* phage Xp10 (390a.a.), *Xanthomonas* phage Xop411 (390a.a.), *Streptomyces* phage Mu1/6 (419a.a.), *Flavobacterium* phage 11b (379a.a.), *Corynebacterium* phage BFK20 (413a.a.), Enterobacteriophage SFV (409a.a.) and *Burkholderia* phage phi1026b (418a.a.). The sequences were analyzed by ClustalW (Appendix B). The residue numbers above the protein sequence correspond to the HK97 gp5 residues. The partial region of HK97 gp5 is shown from residue 160 to 167. The HK97 sequences are underlined written in color blue. The residues corresponding to the HK97 residue V163 are written in color red.



**Figure 19: The close-up view of the HK97 head crosslink site**

a. The close-up view of one HK97 head crosslink site at the hexamer-pentamer-hexamer junction. The residues K169, N356, E363, and V163 are represented in the color blue, red, magenta, and yellow. b. A view with the residues K169, N356, and E363. c. A view with only two crosslinking residues K169 and N356.

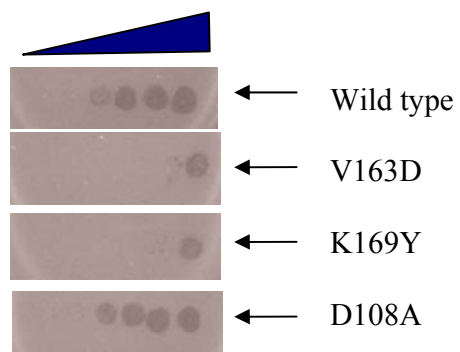


### 4.3 RESULTS

#### **Mutant V163D does not produce functional HK97 gp5**

To test whether a hydrophobic residue is necessary at position 163 in HK97 gp5, mutant gp5 V163D was constructed in plasmid pV0 as pV0-V163D for examining the gp5 functionality. The plasmid pV0-V163D contains a truncated fragment of the wild type portal protein gene, the protease gene, and the V163D mutant capsid protein gene. The mutant V163D was induced in the expression strain BL21(DE3), pLysS, and the gene 5<sup>-</sup> amber phage was spotted on the above bacterial lawn to test the biological activity of the constructed mutant V163D. The results show that mutant V163D is much less (<1000) able to complement the gene 5<sup>-</sup> amber phage, indicating that V163D is a lethal mutation and mutant V163D cannot produce functional gp5 (Figure 20).

**gene 5<sup>-</sup> amber phage**

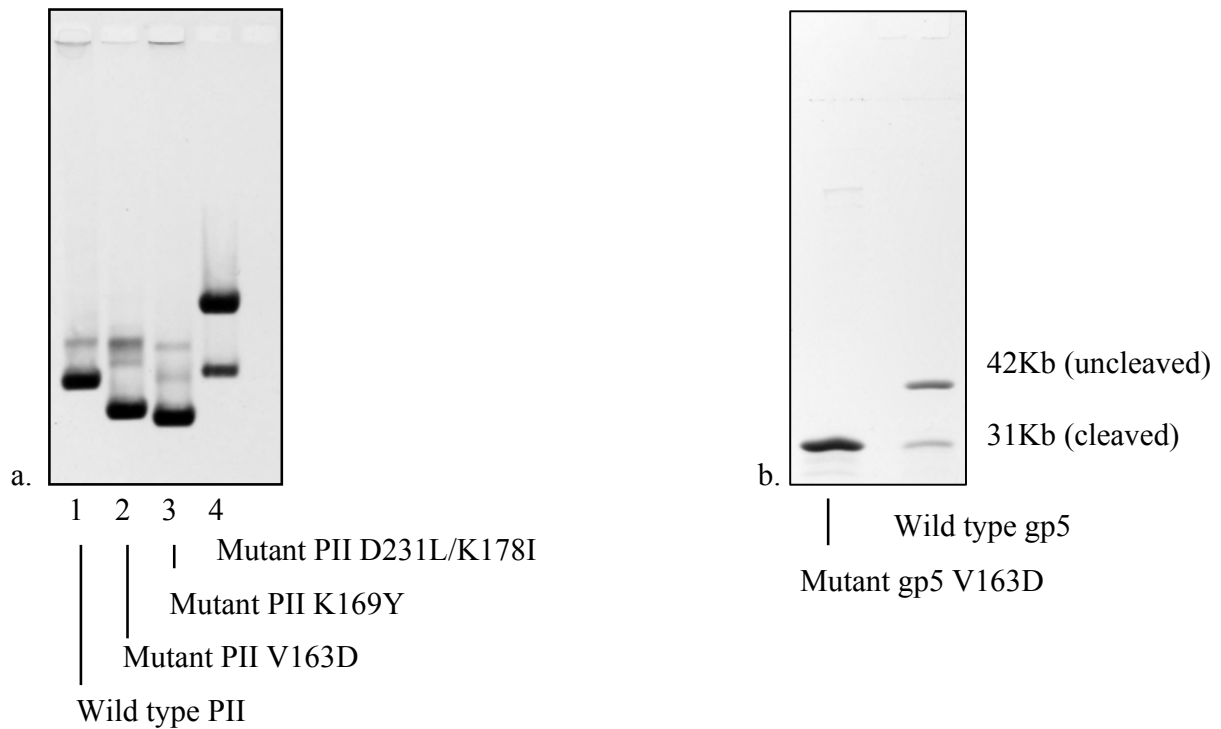


**Figure 20: The mutant gp5 V163D cannot complement the gene 5<sup>-</sup> amber phage**

The wild type gp5, the mutant V163D, K169Y, and D108A, are carried in the plasmid pV0 individually and expressed in the expression strain BL21(DE3), pLysS. A ten-fold serially dilution of the gene 5<sup>-</sup> amber phage was spotted on the bacterial lawns. The mutant K169Y and D108A are negative and positive controls, respectively.

### **The mutant gp5 V163D can assemble Prohead II**

Unlike L361 and M339, residue V163 is located on the E loop and is not buried in the folded gp5 structure until the E loop has descended onto the neighboring subunit prior to crosslinking (Figure 17, Appendix C). Thus, it may be possible that a single substitution at residue 163 might not cause a defect in gp5 folding or assembly prior to the stage of expention and crosslinking. To test whether the mutant V163D assembles any prohead, mutant V163D and the protease gp4 were co-expressed in *E. coli*. Proheads were produced and were determined to be Prohead II based on the cleaved gp5 (31kD) on an SDS PAGE (Figure 21). Prohead II V163D can be purified *in vitro*, suggesting that mutant V163D can be cleaved by gp4, producing Prohead II (Figure 21). Besides, Prohead II V163D migrated differently from the wild type Prohead II on the agarose gel (Figure 21), consistent with the expectation that the net surface charge of the Prohead II V163D is different from that of the wild type Prohead II since residue 163 is not buried inside the structure. All of the above data show that mutant V163D can produce gp5, assemble capsomers, and assemble precursor capsid that undergo proteolysis to yield Prohead II, suggesting that this mutation does not cause a general defect in gp5 folding.

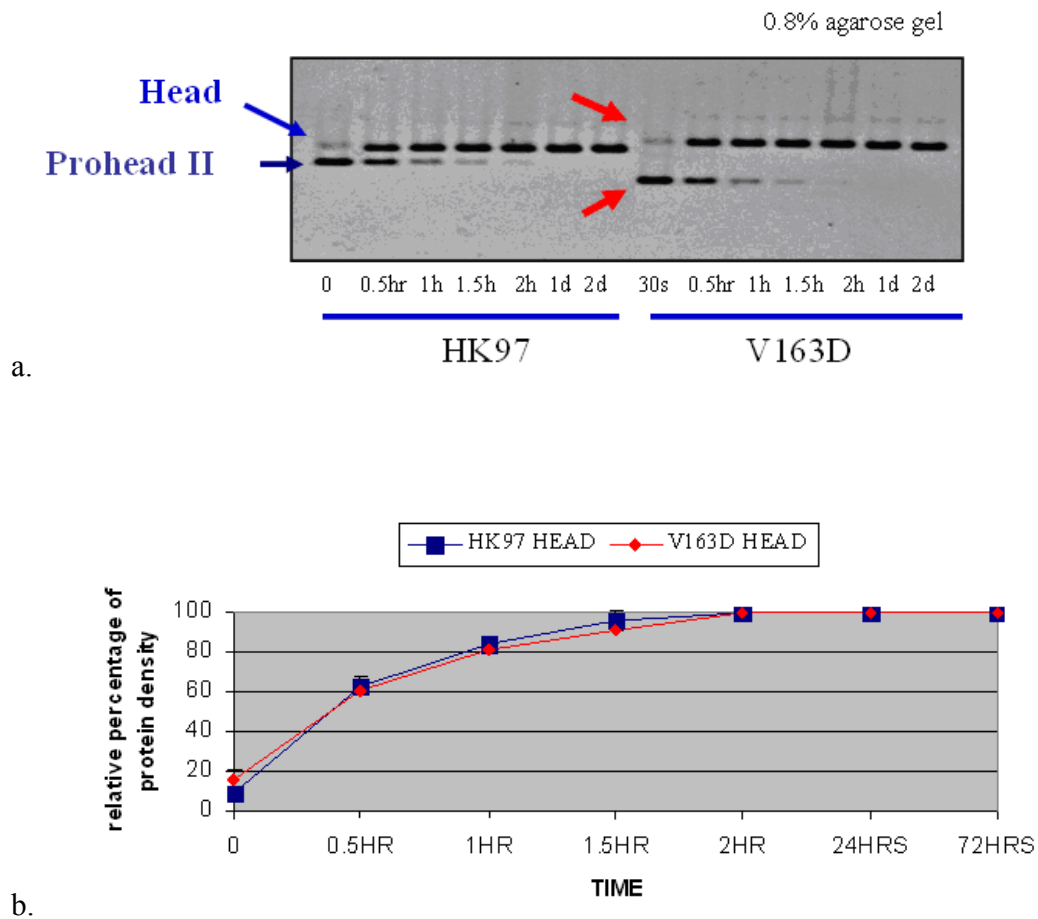


**Figure 21: Prohead II V163D can be purified *in vitro***

a. Prohead II V163D was loaded in lane 2 on a 0.8% agarose gel. Lanes 1-4 are the wild type Prohead II, mutant Prohead II V163D, K169Y, and the double mutant D231L/K178I. b. The cleaved mutant gp5 V163D was detected in lane 1 on a 10% SDS PAGE. A mixture of wild type Prohead I and Prohead II was loaded in lane 2 for the controls as the un-cleaved and cleaved gp5. The cleaved gp5 is labeled as 31kD and the uncleaved gp5 is labeled as 42kD.

### **The mutant Prohead II V163D can expand**

The previous data suggest that mutant gp5 V163D has no significant defect in capsid assembly or proteolysis (Figure 21). But the complementation results show that mutant gp5 V163D cannot produce functional gp5 (Figure 20). Therefore, it is expected that lethal mutation V163D may cause certain defects in capsid maturation. There are two dramatic conformational changes during HK97 capsid maturation: prohead expansion and head protein crosslinking. The expansion ability of Prohead II V163D can be examined *in vitro*. The wild type and V163D Prohead II were treated with the expansion buffer—high salt/low pH buffer—and collected for examining on an agarose gel. The results show that Prohead II V163D does expand and has no significant difference in expansion kinetics between the mutant V163D and the wild type under the experimental conditions (Figure 22), indicating that Prohead II V163D is proficient for HK97 expansion. To further confirm the above results, the expansion was induced with DMF expansion buffer *in vitro*. The results show that Prohead II V163D expands (data not shown), consistent with the above data.

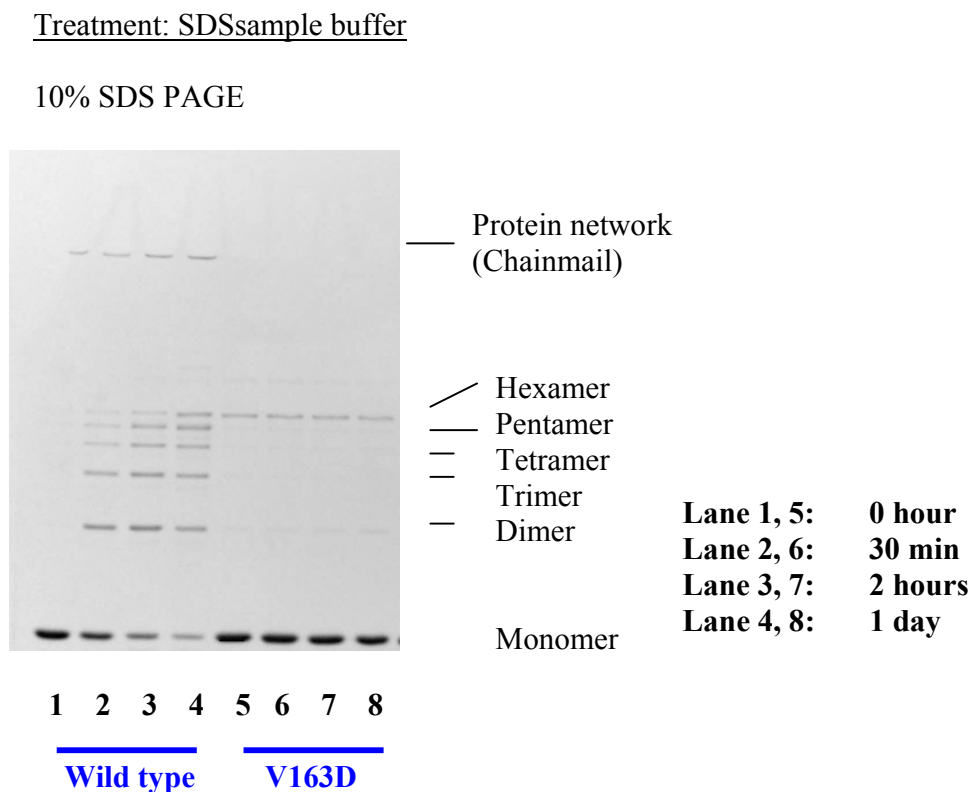


**Figure 22: Prohead II V163D can expand *in vitro***

a. HK97 Prohead II V163D and the wild type Prohead II were treated with the high salt/low pH expansion buffer for 30min, 60min, 90min, 120min, 1 day, and 2 days and were analyzed on a 0.8% agarose gel. Prohead II V163D migrated differently from the wild type Prohead II. b. The graph shows the changes in the measured intensity of the expanded prohead band.

### The mutant Prohead II V163D is defective in head protein crosslinking

In order to determine whether the single substitution V163D causes any defect in HK97 crosslinking, Prohead II V163D and wild type Prohead II were incubated individually with SDS sample buffer to induce crosslinking *in vitro*. The samples were induced for 30min, 2 hours, and 1 day and were collected to examine on an SDS PAGE. The results show that V163D cannot crosslink, but wild type Prohead II crosslinks after 1 hour incubation, indicating that V163D is defective in head protein crosslinking (Figure 23).



**Figure 23: Prohead II V163D cannot fully crosslink *in vitro* (treatment: SDS sample buffer)**

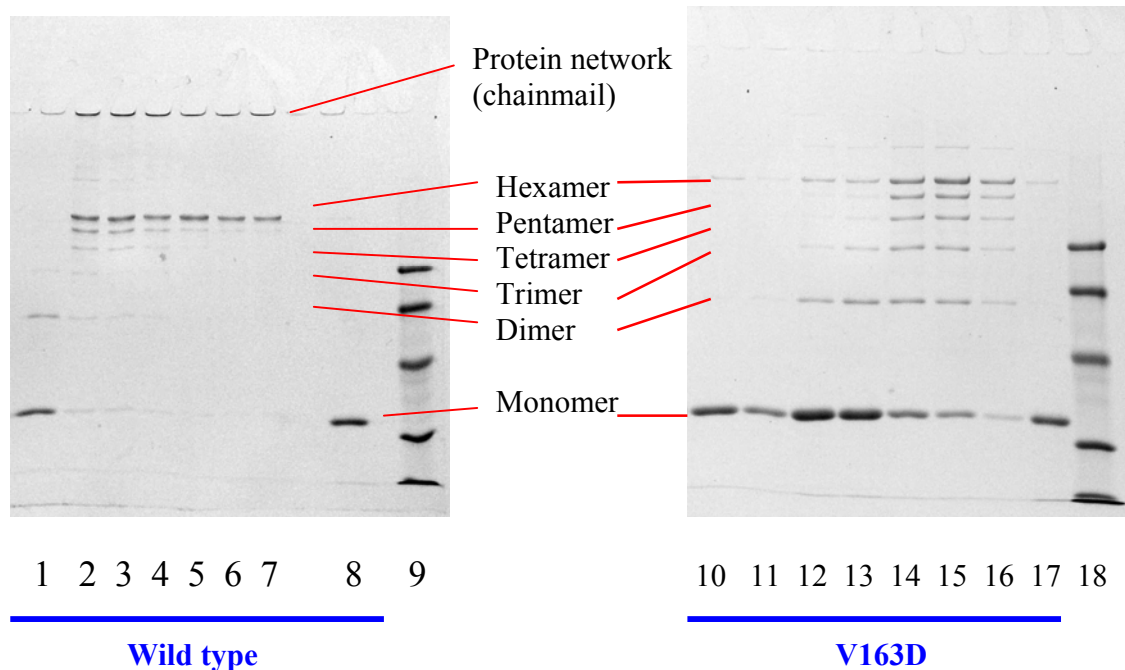
The Prohead II V163D and the wild type Prohead II were treated with SDS sample buffer for 0 min, 30 min, 2 hours, and 1 day and were analyzed on a 10% SDS PAGE.

To further examine the crosslinking ability of Prohead II V163D *in vitro*, and to confirm the previous results, the Prohead II V163D was examined for the crosslinking ability under two other conditions. First, Prohead II V163D and wild type Prohead II were incubated individually with high salt/low pH buffer for one hour, and samples were neutralized with Tris-Cl (pH 8.0) to induce the crosslinking reaction. The results show that V163D only has a few crosslinking bands: dimer, trimer, and hexamer, but no sign of large crosslink products or chainmail network on an SDS PAGE (Figure 24).



Treatment: high salt/low pH buffer

10% SDS PAGE



**Lane 1, 10: 0 min**  
**Lane 2, 11: 30 min**  
**Lane 3, 12: 60 min**  
**Lane 4, 13: 120 min**  
**Lane 5, 14: 1 day**  
**Lane 6, 15: 4 days**  
**Lane 7, 16: 7 days**  
**Lane 8, 17: un- neutralized Prohead II wild type and V163D**  
**Lane 9, 18: protein ladders**

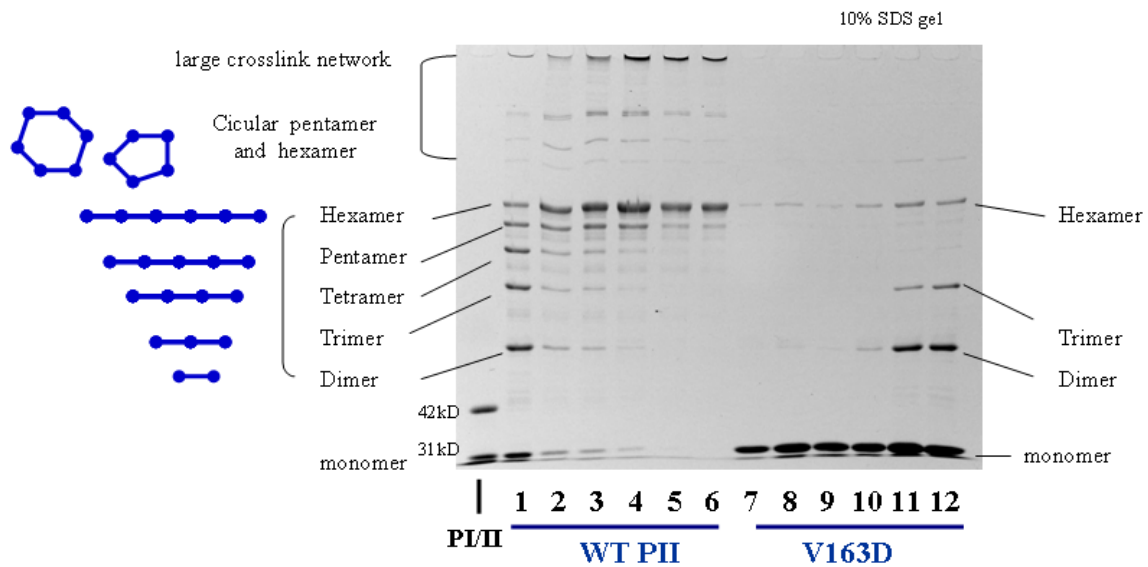
**Figure 24: Prohead II V163D cannot fully crosslink *in vitro* (treatment: high salt/ low pH buffer)**

The Prohead II V163D and wild type Prohead II were treated with high salt/low pH buffer for 1 hour and neutralized to induce the crosslinking reaction. Lanes 1-7 are the wild type samples collected at 0 min, 30 min, 60 min, 120 min, 1day, 4day, and 7day. Lanes 10-16 are the mutant samples collected at 0 min, 30 min, 60 min, 120 min, 1day, 4day, and 7day. Un-treated wild type and mutant samples are in Lanes 8 and 17. The same protein ladders are in Lanes 9 and 18.

Similar results were obtained when samples were treated with DMF buffer. Prohead II V163D did not have any sign of large crosslink products or chainmail network, but wild type Prohead II crosslinks immediately after neutralization (Figure 25). To summarize, Prohead II V163D shows substantial crosslinking defects, indicating that V163D defects in the crosslinking is a consistent phenomenon. V163D has no significant defect in the subunit gp5 folding, prohead assembly, and prohead expansion, I conclude that V163D specifically blocks the HK97 head protein crosslinking and residue 163 plays an essential role in the HK97 head protein crosslinking.

Treatment: DMF buffer

10% SDS PAGE



**PI/PII:** a mixture of wild type Prohead I and Prohead II as the control for un-cleaved (42kD) and cleaved (31kD) gp5

**Lane 1, 7:** 0 min

**Lane 2, 8:** 30 min

**Lane 3, 9:** 60 min

**Lane 4, 10:** 180 min

**Lane 5, 11:** 1 day

**Lane 6, 12:** 2 days

**Figure 25: Prohead II V163D cannot fully crosslink *in vitro* (treatment: DMF buffer)**

The Prohead II V163D and the wild type were treated with the DMF buffer for 1 hour and were neutralized to induce the crosslinking reaction. Lanes 1 to 8 are the wild type collected at 0 min, 30min, 60min, 180 min, 1 day, and 2 days. Lanes 9 to 14 are the V163D collected at 0 min, 30min, 60min, 180 min, 1 day, and 2 days. The mixture of the wild type Prohead I and Prohead II were loaded for the controls as the uncleaved gp5 and cleaved gp5. All samples were analyzed on a 10% SDS PAGE.

### **The first-site suppressor L163 of gp5 is sufficient to provide viable HK97**

In order to understand the local protein network around residue 163, the suppressors of mutant lysogen DT-V163D, which contains a single substitution V163D at residue 163, were isolated. The linear dsDNA fragment was amplified from the plasmid pV0-V163D for recombineering into the HK97 lysogen (see Chapter 3).

DT-V163D was induced by UV light and four revertants were identified. All of the revertants contained the designed silent mutation in gp5, indicating that these revertants contain reversion mutations, and are not contaminating plasmid or wild type phage HK97. Revertants grew well at 20°C, 30°C, 37°C, and 42°C, suggesting that they are temperature-independent. The range of the mutation frequency is one in  $10^9$ - $10^{10}$ . Three of four revertants are true revertants (Table 6). True revertants carry V163V (GTT) at residue 163, suggesting that DT-V163D more readily mutates back to the wild type amino acid instead of obtaining second-site suppressors. The mutation V163V (GTT) may be favored by the fact that it entails only a single nucleotide change from the V163D mutant (GAT) (Table 6). The only pseudo-revertant carries V163L (CTT) at residue 163, showing that a different hydrophobic residue Leu is sufficient to produce viable bacteriophage HK97 (Table 6).

**Table 6: The tolerated amino acids at gp5 residue 163 found in HK97 revertants**

Four plaques were isolated from the *E. coli* lawn Ymel. Three of them are true revertants and one of them is a pseudo-revertant. The first column lists the original residue 163. The second column lists the mutations at residue 163 in the constructed lysogens. The third column lists the identified mutations at residue 163 in the isolated revertants.

Residue	Mutation in mutant lysogen	mutation in revertant
V163 (GTG)	V163D (GAT)	V163L (CTT)
V163 (GTG)	V163D (GAT)	V163V (GTT)
V163 (GTG)	V163D (GAT)	V163V (GTT)
V163 (GTG)	V163D (GAT)	V163V (GTT)

### **The residues Val, Leu, Thr, Cys, and Ile can be tolerated at position 163 in gp5**

Construction of the nineteen (other than Val) possible substitutions at residue 163 in the prophage's gene 5 is a laborious work. To have a better way to examine what other residues can be tolerated at position 163, the mutant lysogens, were constructed and induced for isolating functional phages by randomly introducing different amino acids at residue 163 (see Chapter 2, Method).

A population of linear dsDNA gene 5, carrying a quasi-random mixture of codons at position 163, was recombined into lysogen HK97 to generate a mutant library at residue 163. Two mutational results can occur when the phages are isolated. One is that the phage has a tolerated residue at the gp5 position 163. The other is the phage retains a lethal mutation at residue 163 but has a substitution to suppress the lethal mutation. The second event would probably be very rare. I have only observed the first event in my results.

The mixed population of mutant lysogens was induced by UV light immediately after processing the homologous recombination. The phages with a tolerated residue at 163 can be isolated since they are able to produce a plaque on the *E. coli* lawn Ymel. The randomly recombined gp5 with various amino acids at the position 163 provide a mutagenesis *in vivo*. Since the residue 163 is essential for functional gp5, the phage HK97 viability offers a selection. Twenty-nine survivors were obtained from two individual experiments. The plaques were randomly chosen from the *E. coli* lawn Ymel. All of the survivors carry the designed silent mutation next to the residue 163, indicating that they are not contaminants. The most common residue at 163 found in the phages (13 out of 29) is Val (GTA), implying that most of the survivors revert to the wild type amino acid. The other sixteen phages carry Thr, Ile, or Cys. Nine out of total sixteen survivors have Thr (ACT or ACG), which is the second most abundant

residue found at position 163 in the collection. Ile (ATT) and Cys (TGT) are the other two tolerated residues. These data show that amino acids Val, Thr, Cys, or Ile are functional at residue 163 to provide viable HK97.

**Table 7: The amino acids Val, Thr, Cys, and Ile at residue 163 were isolated in the HK97 survivors**

Total of 29 survivors were isolated from two individual experiments. The first column shows the amino acids that were identified at gp5 residue 163. The second column shows the number of the isolated plaques. The third lane shows the codons for the tolerated amino acids.

Residue at 163	# plaques of total	Codon
<b>Val</b>	<b>13</b>	GTA or GTG
<b>Thr</b>	<b>9</b>	ACT or ACG
<b>Cys</b>	<b>4</b>	TGT
<b>Ile</b>	<b>3</b>	ATT or ATA

## 4.4 SUMMARY & DISCUSSION

### Characterization of the gp5 residue 163

Experimental evidence of the capsid crosslink was found in several dsDNA bacteriophages *e.g.* D3, XP10, BFK20, Φ11b (Borriss et al., 2007; Bukovska et al., 2006; Gilakjan and Kropinski, 1999; Yuzenkova et al., 2003). The three crosslinking residues Lys, Asn, and Glu are conserved in these potential homologs of the major capsid protein (Borriss et al., 2007; Bukovska et al., 2006; Gilakjan and Kropinski, 1999; Yuzenkova et al., 2003). Interestingly, the residue V163 is also conserved among them (Figure 18, Appendix B). This suggests that residue 163 not only provides a crucial role in the HK97 crosslinking, but that it may serve a common role among dsDNA bacteriophages that form capsid crosslinks.

The complementation results showed that mutant V163D cannot produce functional gp5 to produce HK97, implying that a single substitution V163D in the gp5 can cause a functional defect (Figure 20). Prohead II V163D can be detected and purified *in vitro*, showing that V163D does not cause any defect in the prohead assembly and the proteolysis (Figure 21). Interestingly, the crosslinking results showed that mutant V163D was specifically defective in head protein crosslinking, suggesting that residue 163 plays a functional role in the HK97 crosslink formation (Figure 23, 24, and 25). Prohead II V163D expands as efficiently as the wild type, indicating that it functions normally during capsid maturation except for the defect in crosslinking (Figure 22).

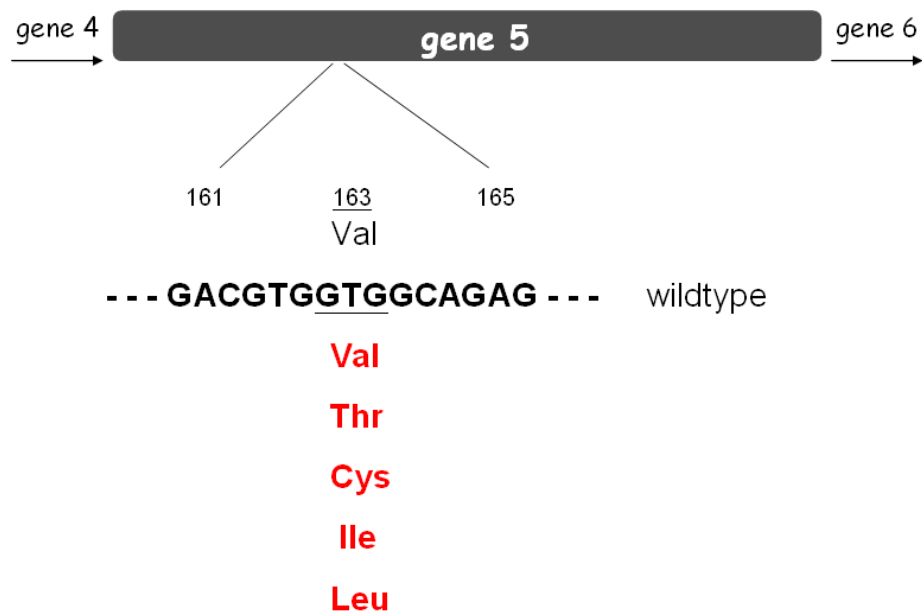
The wild type HK97 crosslinking produces a ladder of the crosslinking bands and some large chainmail proteins on an SDS PAGE when incubated with various chemicals (Figure 23,



24, and 25). The crosslinking ability of Prohead II V163D has been examined *in vitro* using three different treatments: 1) SDS sample buffer, 2) high salt/low pH buffer (pH 4), 3) 36% DMF (pH 8) buffer. Prohead II V163D produces a small amount of crosslinking bands on an SDS PAGE after incubation with SDS sample buffer overnight, while the wild type Prohead II can crosslink nearly to completion after the an hour of incubation (Figure 23). Few, if any, crosslinking bands have been seen when samples were treated with either high salt/low pH buffer or DMF buffer (Figure 23, 24, and 25). No large crosslinking protein networks (chainmail) showed up even after a four week exposure (data not shown). These data together show that even some crosslinking can occur *in vitro* for the V163D mutant, it occurs much more slowly than in wild type and never goes to completion under the conditions examined. Rapid and complete crosslinking is required for producing viable phage HK97 (Dierkes et al., 2009), and the failure of the mutant to support phage growth is plausibly explained by the crosslinking defect.

These previous results together demonstrated that Val, Leu, Thr, Cys, or Ile is able to support efficient crosslink formation when they are at residue 163. These are all non-charged (polar or non-polar) amino acids, and the one charged amino acid that was examined, Asp in mutant V163D, did not support efficient crosslinking. There is no aromatic amino acid nor any amino acid with a small side chain found at residue 163 in the survivors, suggesting that only the residues with a large enough side chain may fit at residue 163 to provide wild type gp5 function. Furthermore, the genetic results revealed that amino acids Val, Leu, Thr, Ile, or Cys at position 163 can support phage growth, and by inference, allow efficient crosslinking (Figure 26). No charged amino acid was isolated, which agrees with the previous results that a negatively charged residue at position 163, *i.e.* V163D, cannot support efficient head protein crosslinking. Four of five tolerated residues, Val, Leu, Ile, and Cys (Nagano, Ota, and Nishikawa, 1999) are

hydrophobic, but Thr is not. This would appear to argue against our previous hypothesis that a hydrophobic residue is required at residue 163.



**Figure 26: The tolerated amino acids at gp5 residue 163 in the isolated HK97 survivors**

Codons 161 to 165 of the wild type HK97 gene 5 nucleotide sequences are shown. The tolerated residues are listed in color red.

To better understand the genetic results, the side chains of five residues were compared side by side. It was found that Val and Thr have similar side chains among amino acids. The difference between them is that Val has a methyl group (-CH<sub>3</sub>), not the hydroxyl group (-OH) which is carried in Thr. It may be instructive in this context to consider the fact that no serine substitutions at position 163 were recovered. Threonine differs from serine only in having a methyl group attached to the beta carbon, which is a methylene carbon in serine. Possibly the extra methyl group in threonine provides enough hydrophobic character to support crosslinking. Furthermore, all of the five tolerated residues have relatively similar side chain lengths, *i.e.* not too large like Trp, or not too small like Gly, indicating that the size of the residue 163 may be a concern for fitting the residue 163 into the crosslink site.

Residue 163 is located at the end of the gp5 E loop (Figure 17, 19 and 27, Appendix C). The E loop protrudes out of the Prohead II shell surface but lies on the head surface after capsid maturation. One of the crosslinking residues, K169, and the residue V163 are both located in the E loop and are translocated close to the other crosslinking residue N356 during the conformational rearrangements of shell expansion (Figure 17, 19, and 27, Appendix C). After translocation, residue V163 is located very close to the crosslinking central area and faces toward the crosslink site (Figure 19 and 27, Appendix C). It is reasonable that residue 163 requires a specific size of side chain to fit into the crosslink site. Furthermore, V163D cannot produce viable HK97, so the other negative charged residue, Glu, propably cannot be tolerated either.

Here I demonstrat that the size and the character of residue 163 may affect the HK97 crosslinking, and the precise structure and hydrophobic characteristics are both important for the functions of residue 163. Prohead II V163D, carrying a charge at its side chain, may remain at an unproper position and cannot be buried inside the crosslink pocket. The crosslink defect of the

mutant V163D may be caused by the failure of E loop translocation and the failure to enclose the residue 163 accurately into the crosslink site.

### **Genetic analysis of gp5 residue V163 *in vivo***

The primer WNNCACATCGGCGTTATTGGTAAACACC, which was designed to randomly introduce amino acids in residue 163 (NNW) (Chapter 2, Method), are able to provide Phe, Leu, Ile, Val, Ser, Pro, Thr, Ala, Tyr, His, Gln, Asn, Lys, Asp, Glu, Cys, Arg, and Gly in the lysogen HK97. Thus, the survivors were selected from these constructed lysogens. However, two hydrophobic residues, Met (ATG) and Trp (TGG), were not constructed in the lysogens. In order to test the functional abilities of Met and Trp at residue 163 and to confirm the genetic data, the other designed primer SNNCACATCGGCGTTATTGGTAAACACC can be used to introduce 20 different amino acids (NNS) in lysogen HK97. The survivors can be selected for determine the tolerated residues at 163. The results of this future experiment are expected to be similar to the results in Table 7.

### **The hydrophobic crosslink environment**

The crosslink active site is a region which is packed by two adjacent capsomers (Figure 19 and 25). Two polypeptide chains provide a crosslinking donor residue, N356, and a catalytic residue, E363, respectively (Figure 19 and 27). The crosslinking acceptor residue, K169, and the hydrophobic residue, V163, are located on the E loop on the third polypeptide chain, which caps the crosslink active site (Figure 19 and 27, Appendix C). The residues L361 and M339 are located at the two sides of residue E363 on the same subunit, and residue V163 and L361 flank the crosslinking residue K169 (Figure 17). This suggests that the crosslink active site is packed as a hydrophobic environment where crosslinking reaction occurs.

My results support the model of the crosslink active site as follows: when the lid of the crosslink active site, the E loop, lies close to the other two polypeptides, a pre-crosslinking hydrophobic environment is formed by the surrounding residues, *e.g.* V163, to establish a cavity for burying the crosslinking reaction (Figure 19 and 27, Appendix C). Thus, the residue 163 may be part of an essential environment, *e.g.* hydrophobic pocket, to initiate the protein crosslinking reaction.

The Gram positive bacterial pili are the long strings formed by the crosslinking pilin subunits, which are covalently attached end-to-end by isopeptide bonds (Falker et al., 2008; Ton-That and Schneewind, 2004). This intermolecular crosslink is catalyzed by enzymes called sortases (Falker et al., 2008; Ton-That and Schneewind, 2004). Recently, it was found that the pilin subunits have two distinct intramolecular isopeptide bonds in their subunit N domain and C domain (Kang and Baker, 2009; Kang et al., 2007). Three conserved crosslinking residues were found in the pili of the *Streptococcus pyogenes*, as the residues K36, N168, and E117 in the protein N domain and the residues K179, N303, and E258 in the C domain (Kang et al., 2007).

This indicates that bacterial pili may have chemically similar crosslinking mechanism to the phage HK97. In addition to the *S. pyogenes*, the pili K-N crosslink was also found in some other Gram positive bacterial pili, *e.g.* the pilin protein SpaA in the *C. diphtheriae* (Kang et al., 2009), the pilin protein BcpA in the *B. cereus* (Budzik et al., 2008), and the pilin protein GBS52 in the *S. agalactiae* (Kang et al., 2007). These findings suggest that the HK97 crosslinking reactions may be a widespread mechanism. The phage capsid and the bacterial pilin may have evolved a similar crosslinking reaction. However, the fact that the two crosslinking residues and the catalytic glutamate of HK97 come from different protein subunits while the corresponding residues in the pilin case are on the same polypeptide chain suggests that this is a case of convergent evolution, that is, independent evolution of a similar mechanism in the two cases.

Interestingly, some of the hydrophobic residues were found to cluster around the two intramolecular isopeptide bonds in the *S. pyogenes* pili (Kang and Baker, 2009). For instance, Phe54, Ile56, and Phe98 are close to the crosslinking residues K36 and N168 in the protein N domain, suggesting a hydrophobic environment around the pilin crosslink (Kang and Baker, 2009). These results from the Gram positive bacteria are consistent with our hypothesis for the hydrophobic crosslink environment: hydrophobic residues, *e.g.* V163, around the HK97 K-N isopeptide bonds are required to form the crosslink active site.

Hydrophobic residues were shown to be required for other types of crosslinking reactions: elastin matrix formation. Three hydrophobic domains flanking two crosslinking domains for the recombinant human elastin polypeptides are sufficient for the elastin self-assembly (Bellingham et al., 2003). These polypeptides can spontaneously self-reorganize into fibril-like structures, which is indistinguishable from that of the native elastin (Bellingham et al., 2003). The hydrophobic domains are necessary for the above self-aggregation, suggesting that the residues

in the hydrophobic domains may contribute to produce covalent intermolecular crosslinking or stabilize the insoluble matrix (Bellingham et al., 2003). It is interesting to note that the reaction does not need any enzyme catalysis. The details of this mechanism for assembling the above matrix remain unknown. But it is clear that the hydrophobic domains are required for the self-assembling elastin.



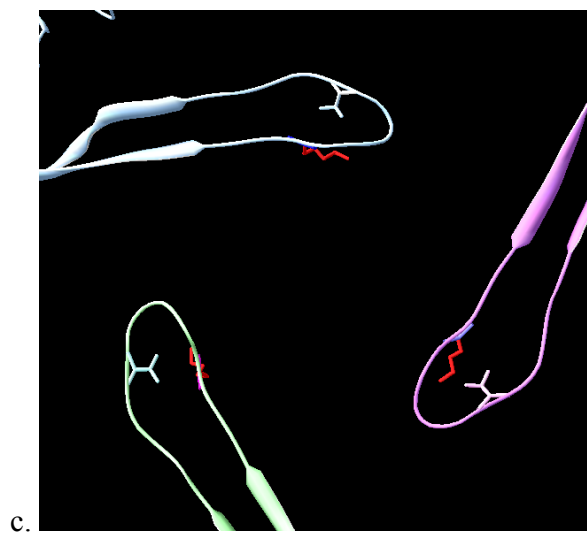
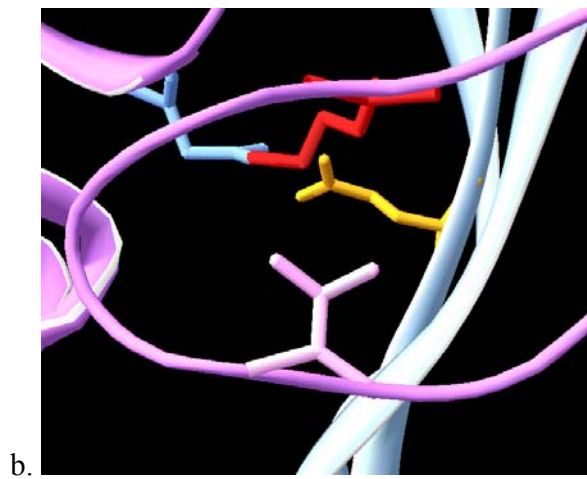
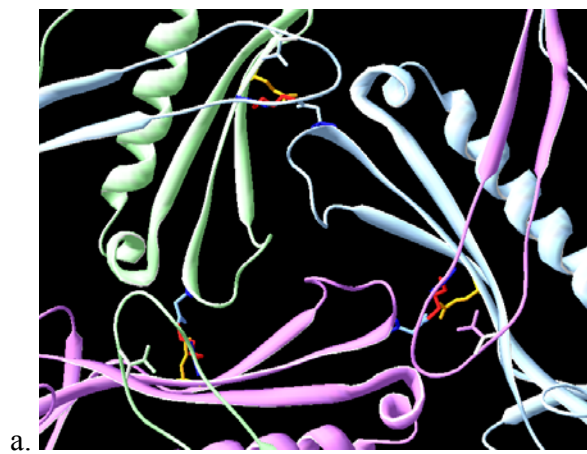
### **The mutant V163D is defective in head protein crosslinking**

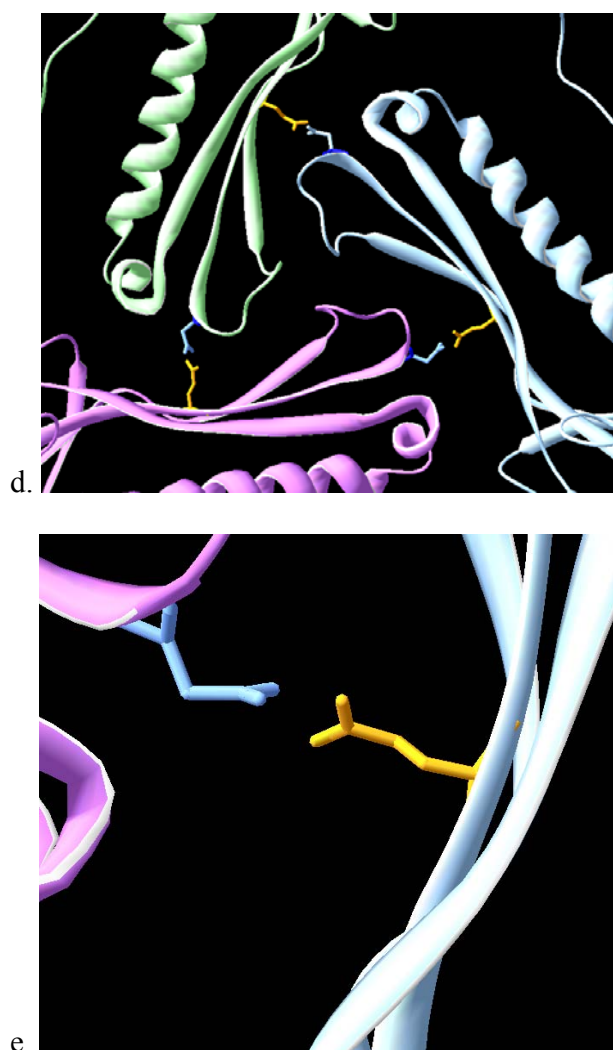
The head protein self-crosslinking is nearly the last step of HK97 capsid maturation, producing the covalent K-N isopeptide bonds between two adjacent gp5's without the aid of any enzyme. How HK97 triggers and catalyzes such crosslinking on its capsid remained unclear. Recently, the catalytic residue 363 has been characterized (Dierkes et al., 2009). It was found that residue 363 is essential for HK97 viability. The single substitutions E363A, E363D, E363H, E363K, E363Y, E363N, and E363Q in gp5 block gp5 functionality and the producing mutant gp5's interfere with the wild type phage growth (Dierkes et al., 2009). The crosslinking abilities of the mutant E363D, E363H, E363Q, E363N, and E363A are differently defective (Dierkes et al., 2009).

Among all the mutants, E363Y cannot assemble any prohead (Dierkes et al., 2009). Except for the mutant E363D, all of the other mutations at residue 363 can accumulate Prohead I but result in incomplete proteolysis (Dierkes et al., 2009). This indicates that some of these mutants are found to have defects before reaching the crosslink. Thus, residue 363 may not only be involved in the crosslinking but also play roles in the early assembly steps. The mutant V163D is different from these E363 mutants. The single substitution V163D specifically affects the crosslinking reaction, which was determined by the *in vitro* crosslinking assays (Figure 23, 24 and 25). Although only one mutation at residue 163, V163D, was examined in this study, it was shown that residue 163 may have more functional specificity than the residue 363 in the head protein crosslinking.

### **The model of the hydrophobic crosslink active site**

Three residues K169, N356, and E363, are far apart from each other in their isolated subunit (Figure 27: their side chains are labeled in red, blue, and yellow). They are located in the subunit interfaces in the individual pentamers and hexamers of the HK97 capsid. The prohead expansion brings together the three required residues which are located in three different polypeptides (Figure 27b: color in blue, pink and the pink with white trimming) to participate in the crosslinking reaction. Both residues K169 and V163 are located in the E loop (Figure 27), which undergoes a translocation during expansion. The other crosslinking residue N356 is near the outer end of the P domain of an adjacent subunit gp5 (Figure 19, Figure 27d and 27e, Appendix C). The residue 363 at the third isopeptide bond (Figure 27b: blue polypeptide) is located at the interface between two capsomers and projects from a  $\beta$ -strand in the P domain of one capsomer (Figure 19, 27d and 27e, Appendix C). After expansion, residue K169, which projects out of the E loop, translocates next to N356 and faces toward the third residue E363 (Figure 19, 27a, 27b, and 27c, Appendix C). The residue V163 is also translocated to the crosslink site after expansion and possibly initiates the hydrophobic crosslink active site (Figure 27a, 27b, and 27c).





**Figure 27: The HK97 crosslink active site**

a. The view of the HK97 head crosslink site (HK97 Head II structure from PDB: 1OHG). The residues K169, N356, and E363 are represented as their side chains with the color in red, blue, and orange. The residues V163 at three different polypeptides are shown in color pink, blue, and green. b. The close-up view of one isopeptide bond. The crosslinking residues K169 and N356 are linked. The residue E363 is located very close to the isopeptide bond. The side chain of the residue V163 faces toward the isopeptide bond. c. The residues K169 and V163 are located on the E loop. Only three of the E loops are shown. d. The E loop was moved to observe the side chains of the residues N356 and E353, which are located in two different polypeptides. e. The close-up picture of d. The crosslinking residue N356 faces the catalytic residue E363.

A crosslinking mechanism is described as follows: The conformational changes, *e.g.* expansion, bring all crosslink-related residues together into the crosslink site. The residue V163 and K169 in the E loop are translocated close to the other residues N356 and E363. The hydrophobic environment may be determined by hydrophobic residues, *e.g.* V163, which is brought inside by E loop translocation. After the pre-crosslinking environment is established, the crosslinking reaction can begin by the proton transfer from Lys (K169) to Glu (E363) and end with forming the irreversible covalent K-N isopeptide bond.

My results strongly support that residue 163 plays a central role to provide a part of the proper crosslink hydrophobic environment. The crosslink site is a packed region and the distances among the crosslinking residues K169 and N356 and the catalytic residue E363 need to be accurate for the reaction. The tolerated residues provide the information that a precise hydrophobic side chain character of residue 163 is required for establishing a pre-crosslink environment. The proper side chains of the hydrophobic residues allow the formation of the crosslink active site which encompasses the crosslinking.

## **5.0 AN ESSENTIAL INTERACTION BETWEEN RESIDUES D231 AND K178 IN HK97 PROHEAD ASSEMBLY**

### **5.1 ABSTRACT**

*Escherichia coli* bacteriophage HK97 assembles its precursor capsid, Prohead I, from 420 copies of major capsid protein gp5. After assembly with the maturation protease gp4, the proheads undergo proteolysis, producing Prohead II which is composed of the cleaved gp5's. Proteolysis is followed by several stages of maturation including shell expansion and crosslinking to form a mature capsid called Head II. Previous studies showed that the first precursor capsid, Prohead I, is assembled from capsomers which are composed of pentamers and hexamers of the major capsid protein. However, the mechanism by which capsomers assemble accurately into proheads remains unknown. Here we report that non-covalent interactions between adjacent capsomers formed by major capsid protein residues Asp231(D231) and Lys178(K178) are critical for the assembly of capsomers into Proheads. Evidence for a direct interaction includes isolation of second-site suppressors K178V, K178I, or K178N that suppress the lethality of the single substitution D231L in gp5, indicating that the non-covalent interaction between residue 231 and 178—a hydrophobic interaction—replaces the original ionic interaction in HK97 revertants. A structural model of HK97 Prohead II illustrates that interactions between residue 231 and 178 occur not within subunits or within capsomers but between two different gp5's located on

adjacent capsomers. Interestingly, the above interaction is not seen on the mature Head II structure. The double mutants, D231L/K178V, D231L/K178I, and D231L/K178N, self-assemble Prohead II with reduced efficiency but without other aberration, indicating that these mutant pairs at residues 231 and 178 only affect early stages of capsid assembly. Unexpectedly, accumulation of Prohead I could not be detected in the three double mutants, suggesting that three double mutant Prohead Is have certain differences from wild type Prohead I. The data support a model that the non-covalent side-chain interactions of residues 231 and 178 play a central and crucial role at the early stage in HK97 capsid assembly.

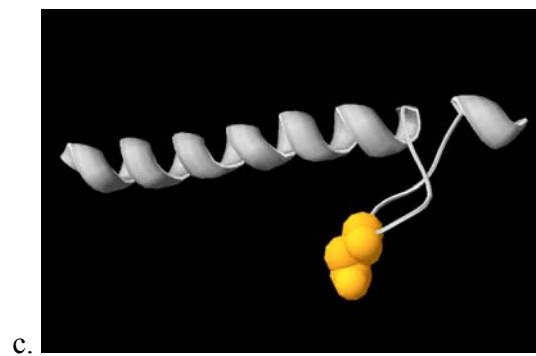
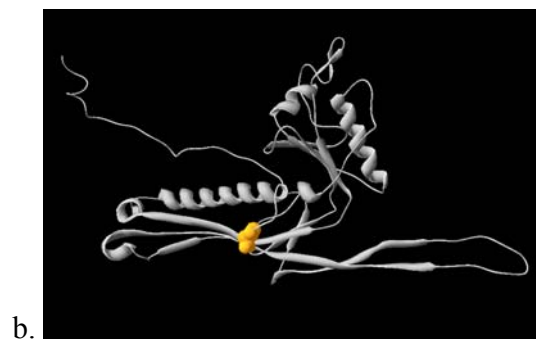
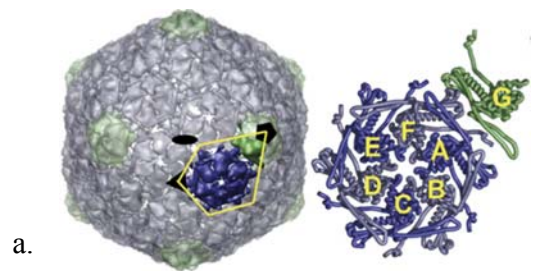
## 5.2 INTRODUCTION

In previous research, the folded subunit gp5 has been described as having four major distinct parts: the N arm (called N terminal), the A domain (called Axial domain), the P domain (called Peripheral domain), and the E loop (called Extended loop), as mentioned in chapter 1 (Figure 3). Here I consider an unusual sub-structure of gp5, designated here as the T loop, containing 9 residues—GDGTGDNLE—located between two alpha helices in the P domain (Figure 3 and 26). Previous data show that functional gp5 is not produced if the entire T loop has been deleted (Y. Li, PhD thesis, University of Pittsburgh), indicating an indispensable role for at least some feature of the loop. The likely importance of the T loop structure in some aspect of gp5 function is reinforced by an amino acid sequence alignment of several related capsid proteins. These sequences align across the HK97 T loop region and across the flanking alpha helical regions with no gaps in the alignment, suggesting conservation of the loop structure (communication in our lab). The amino acid sequences within the putative loops are not well conserved among the capsid proteins in the alignment, but they all have two or more glycines among the loop residues (Appendix B).

In this chapter, I investigate the roles of gp5 residue 231, an aspartic acid (called D231) located at the outer end of the T loop in HK97 Head II. Data reported here show that substitution of residue D231 by either Ala(A) or Leu(L) blocks capsid assembly at the level of preventing capsomers from assembling into prohead shells. Genetic data implicate lysine 178 (called K178) as a crucial interacting partner of D231, and I argue that this interaction is a salt bridge between



capsomers. This non-covalent, side-chain interaction between residues 231 and 178 has a functional role in the initial stage of HK97 capsid shell formation.



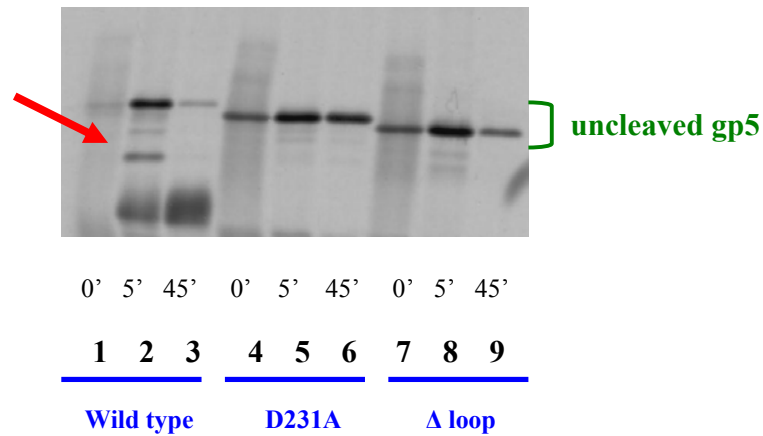
**Figure 28: The T loop, an unusual sub-structure in the gp5 P domain**

a. The three dimensional structure of a single subunit from the HK97 Head II (Gan et al., 2006). b. Residau D231, shown as an orange space-filling model is located at outer end of the gp5 T loop. c. A close-up picture of the long alpha-helix and the T loop in the P domain.

## 5.3 RESULTS

### **The gp5 T loop is essential for gp5 functionality**

To test whether the T loop is essential for gp5 function, mutant gp5  $\Delta$ loop was constructed and tested for the ability to support phage HK97 growth. This mutant gp5 was expressed from the plasmid and tested for its ability to complement a superinfecting phage carrying an amber mutation in gene 5 (called gene 5<sup>-</sup> amber phage) (see Chapter 3, Material). No complementation was observed (data not shown), arguing that the T loop is indispensable for gp5 function. Further studies of the mutant gp5  $\Delta$ loop show that it can produce the mutant subunit gp5 (Figure 29) but produces at most a trace of capsomers and no proheads (Figure 30a), which is as monitored on an SDS PAGE (for gp5), on a native polyacrylamide gel (for capsomers) and on an agarose gel (for capsomers and proheads). The phenotype of the  $\Delta$ loop mutant was not examined further. However, the observation that it appears to be blocked at an earlier stage than are the T loop point mutants described below (*i.e.* before capsomer formation) is consistent with the possibility that  $\Delta$  loop has a folding defect.

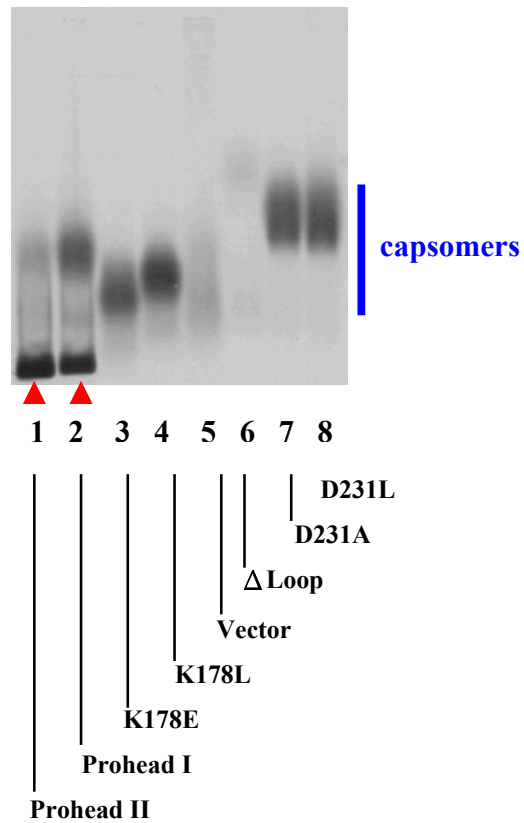


**Figure 29: The gp5 mutants D231A and Δ loop cannot be cleaved by the HK97 protease gp4**

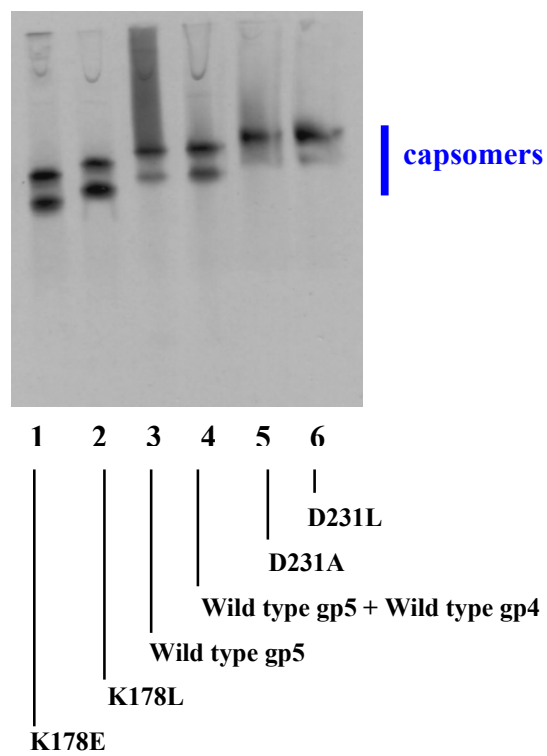
The wild type gp5 (lane 1-3), the single mutant gp5 D231A (lane 4-6), and the deletion mutant gp5 Δloop (lane 7-9) were co-expressed with the protease gp4 individually and radio-labeled with  $^{35}\text{S}$ -Met. All samples were collected after incubating for 0 min, 5 min, and 45 min and were analyzed on a 10% SDS PAGE.. The mutants D231A and Δloop are not cleaved by gp4 while the wild type gp5 can be cleaved by gp4. The cleaved wild type gp5's are pointed by the red arrow.

a.

**Proheads**



b.



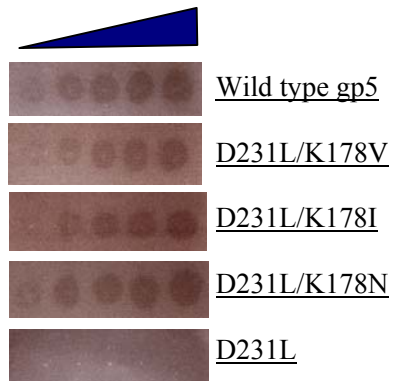
**Figure 30: The gp5 single mutants D231L, D231A, K178E, and K178L assemble capsomers but do not assemble proheads**

a. The single mutant gp5's  $\Delta$ loop, K178E, K178L, D2131L, and D231A were expressed alone and radio-labeled with  $^{35}\text{S}$ -Met. The wild type gp5 were expressed alone (labeled Prohead I) or co-expressed with the protease gp4 (labeled Prohead II) as the Prohead controls. The vector was expressed as the negative control. Lane 1-8 are wild type gp5 w/o gp4, wild type gp5 with gp4, mutant K178E, K178L, vector only, mutant  $\Delta$ loop, D231A, and D231L. Capsomers were detected in lane 3, 4, 6, 7, and 8 while proheads were detected in lane 1 and 2. All samples were analyzed on a 0.8% agarose gel. b. Two types of capsomers, pentamers and hexamers, from single mutant gp5's K178E, K178L, D2131L, and D231A were detected and analyzed on a 7.5% native polyacrylamide gel. The wild type gp5 were expressed alone or co-expressed with the protease gp4 as the controls. Lane 1-6 are mutant K178E and K178L, wild type gp5 w/o gp4, wild type gp5 with gp4, mutant D231A and D231L.

### **The single gp5 mutants D231L and D231A can assemble capsomers but cannot assemble proheads**

The residue D231 is located at the outer end of the T loop in the gp5 structure (Figure 28). To test whether residue D231 is important for gp5 function, the gp5 mutants D231L and D231A were constructed and examined for gp5 function. Both mutants D231A and D231L fail to complement the gene 5<sup>-</sup> amber phage in the spot complementation test (Figure 31 and 38), implying that a single substitution at residue 231 is enough to block the gp5 function. To understand their effects on capsid assembly, the two single mutants D231A and D231L were expressed in *E. coli* and labeled with <sup>35</sup>S-Met. Soluble proteins were collected and examined on an agarose gel (Figure 30a) and on a non-denaturing polyacrylamide gel (Figure 30b). The characteristic broad capsomer bands on the agarose gel are seen for the mutants at levels comparable to those for the wild type, but there are no prohead bands for the mutants. The characteristic two capsomer bands (hexamers and pentamers) are detected on the native polyacrylamide gel. (The prohead does not enter the non-denaturing polyacrylamide gel). These results show that both mutants can assemble capsomers but not proheads, indicating that residue D231 is necessary for assembling capsomers into proheads.

gene 5<sup>-</sup> amber phage

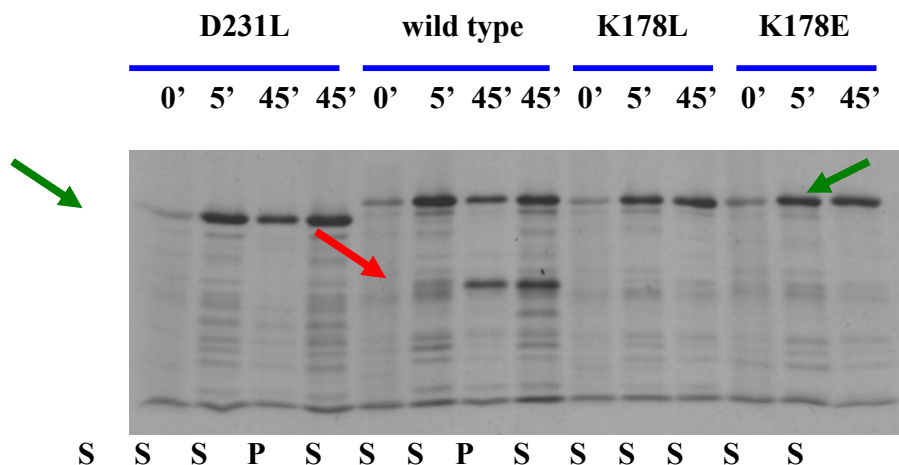


**Figure 31: The double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N can complement the gene 5<sup>-</sup> amber phage**

The double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N were expressed in the expression strain BL21(DE3) individually. A serially diluted gene 5<sup>-</sup> amber phage was spotted on the above bacterial lawns, from right to left. The first row shows that the wild type gp5 complements gene 5<sup>-</sup> amber phage as the positive control. The second, third, and fourth rows show that three double mutants D231L/K178V, D231L/K178I, and D231L/K178N complement the lethality of D231L. The fifth row shows that mutant D231L cannot complement gene 5<sup>-</sup> amber phage as the negative control.



To assay the proteolysis of these two gp5 mutants D231A and D231L, the protein samples were examined on an SDS PAGE. No protease cleavage was detected from mutants D231A and D231L (Figures 29 and 32), suggesting that single mutant gp5's D231A and D231L cannot be cleaved by the protease gp4. Because the site of protease cleavage is distinct in the structure from residue D231, and because in all previous experiments proteolysis has only been observed in cases where capsid assembly can take place, I postulate that the failure of cleavage for these mutants is due to a failure of assembly rather than to an alteration of the protease cleavage site. If this supposition is correct, the results support the conclusion that mutants are defective in assembling capsomers into proheads. It argues against the alternative model that mutant proheads assemble in the cell (at which point proteolysis would be expected) but dissociate to capsomers when they encounter the *in vitro* buffer.



**Figure 32: The gp5 single mutants D231L, K178L, and K178E cannot be cleaved by the HK97 protease gp4**

The single gp5 mutants D231L, K178L, and K178E, and the wild type gp5 were co-expressed with the protease gp4 individually and radio-labeled with  $^{35}\text{S}$ -Met. All samples were collected after incubating for 0 min, 5 min, and 45 min and analyzed on a 10% polyacrylamide SDS PAGE. The wild type gp5 can be cleaved by gp4 while the three single mutants cannot. The supernatant (S) and pellet (P) samples were collected after 45min incubation for mutant D231L and wild type. The cleaved and uncleaved gp5's are pointed by the red and green arrows.

### **Suppressors of D231L map to residue 178**

To understand the function of residue D231, mutant lysogen DT-D231L was constructed with a single substitution D231L in gene 5. DT-D231L was induced by UV light and assayed for production of plaque-forming phages. The same volume of sample without UV light exposure was saved for the untreated control. No plaques were isolated on the control plate, but a small number of plaques were recovered from the induced culture and confirmed to be pseudo-revertants which still carry the single mutation D231L. The average reversion frequency is one in  $10^9$ : there is one revertant plaque for every  $10^9$  plaques produced in a parallel induction of a lysogen with a wild type prophage. A total of seven plaques were chosen from three independent experiments. All of the revertants contain the designed silent mutation in gp5, indicating that these revertants were derived from the mutant lysogens and were not the result of contamination by the wild type phage HK97 or a wild type lysogen. The entire gene 5 in the revertants was sequenced to detect any intragenic suppressor.

Second-site mutations were found in all of the revertants, and all of these mutations were mapped to residue 178, which is a lysine in the wild type gp5 sequence. No other mutation was found in gene 5 except for three intragenic suppressors at residue 178 (Table 8). The same suppressors were isolated from multiple experiments. Two revertants contain the mutation K178V (GTA) and three revertants contain K178I (ATA) (Table 8). The other two revertants contain the K178N but with different codons, AAT and AAC (Table 8). These genetic data strongly suggest a particular relationship between residue 231 and 178, *i.e.* residue 231 may interact with residue 178. The results also suggest that the hydrophobic interactions, D231L/K178V and D231L/K178I, between residue 231 and 178 in two of the revertants can replace a presumed ionic interaction D231/K178 in the wild type HK97 capsid.

**Table 8: The second-site suppressors of lethal mutation D231L in gp5 found in HK97 revertants**

Seven plaques were randomly isolated from the *E. coli* lawn Ymel, and all of them are pseudo-revertants. The first column represents the original residue 231 in gp5. The second column lists the single mutation D231L in the constructed lysogens. The third column lists the single mutation D231L and the second-site suppressors at residue 178 in the isolated revertants.

Residues in the wild type gp5	Mutant gp5 in Lysogen	Mutations in the revertant gp5
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178V(GTA)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178V(GTA)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178I(ATA)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178I(ATA)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178I(ATA)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178N(AAT)
D231(GAT), K178(AAA)	D231L(CTC)	D231L(CTC), K178N(AAC)

### **The identified single mutations K178N, K178I, and K178V suppress the lethal mutation D231L**

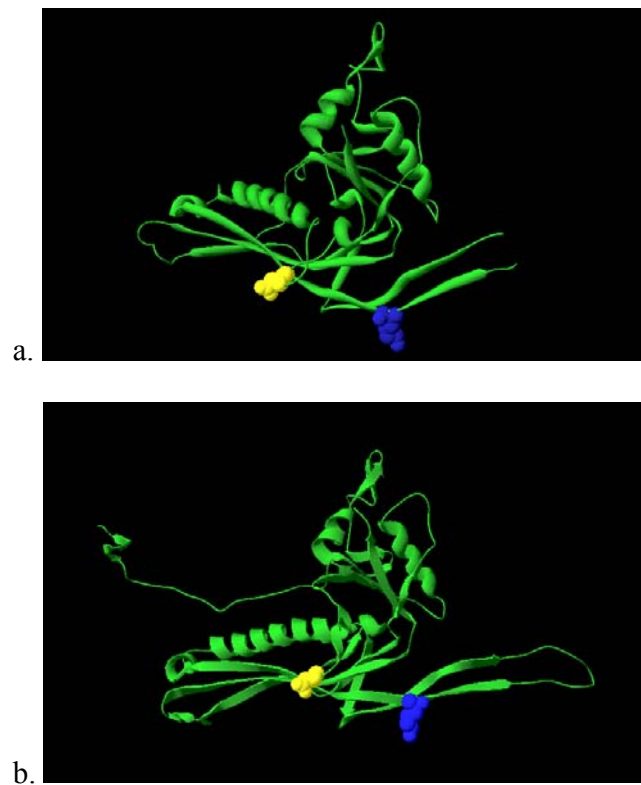
Genetic data presented above gives strong evidence of the relationship between residues 231 and 178. However, the sequences of the revertant genomes have not been determined outside of gene 5, so the possibility that undetected suppressor mutation(s) outside of gene 5 rather than the changes at residue 178 are responsible for suppressing D231L cannot be ruled out. In order to confirm that mutations K178V, K178I, and K178N are the major suppressors of single mutation D231L, three double mutants D231L/K178V, D231L/K178I, and D231L/K178N were constructed in the plasmid pV0 for determining their gp5 functionality. The results, shown in Figure 31, confirm that mutant D231L cannot produce functional gp5 to complement the gene 5<sup>-</sup> amber phage, but all three double mutants, D231L/K178V, D231L/K178I, and D231L/K178N, give full complementation, showing that the single substitutions K178V, K178I, and K178N are sufficient to suppress the lethality of D231L.

### **The gp5 residues 231 and 178 provide an essential interaction to assemble the HK97 capsomers into proheads**

The above data provide evidence of a biological importance of residues 231 and 178, but the exact function of their relationship is unknown. At the beginning of my research, residue D231 was predicted to form a salt bridge to interact with the residue K184 within gp5 to stabilize the gp5 folding (Y. Li, PhD thesis, University of Pittsburgh). Surprisingly, instead of residue K184, residue D231 was found to form a salt bridge with residue K178 of the adjacent gp5 at a nearby capsomer to assemble the capsomers into proheads.

The residue 231 is located in the T loop and residue 178 is located in the E loop (Figure 33). The residue K178 is located far from D231 within a gp5 ( $>26\text{\AA}$ ) and face toward another different direction other than that of residue D231 (Figure 33 and 34), suggesting that residue K178 does not interact with D231 within gp5 or is not involved in the gp5 folding. However, the relationship between them became clear when the locations of residues 231 and 178 were examined on the HK97 Prohead II (Figure 34, and Appendix D) (Gertsman et al., 2009). A structural view of Prohead II includes one pentamer and five hexamers was used (Figure 34a, Appendix D). The pentamer located at the central position is surrounded by five hexamers (Figure 34). Instead of getting buried inside of capsomers, residue 231 and residue 178 face out of each capsomer (Figure 34a, 34b, 34c, and Appendix D). There are five residues 231 and five residues 178 in one pentamer, and there are six residues 231 and six residues 178 in one hexamer (Figure 34, Appendix D). It was found that residue 231 in one gp5 of a capsomer can interact with residue 178 of an adjacent gp5 in a nearby capsomer whereas two residues cannot interact with each other in the same capsomer (Figure 34 and Appendix D). The interactions of residue 231 and 178 were observed either between pentamers and hexamers, or between hexamers and

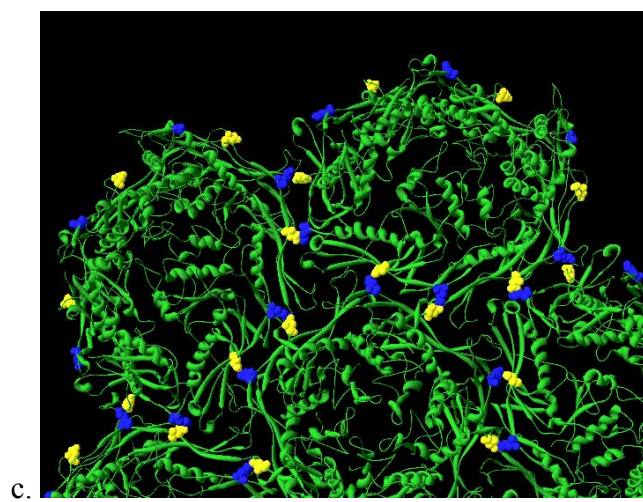
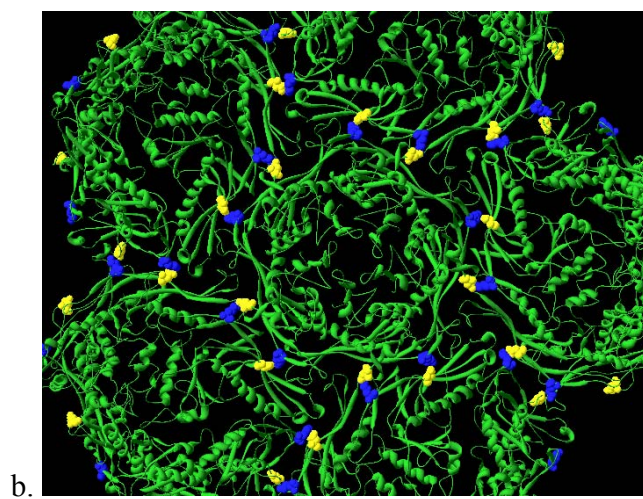
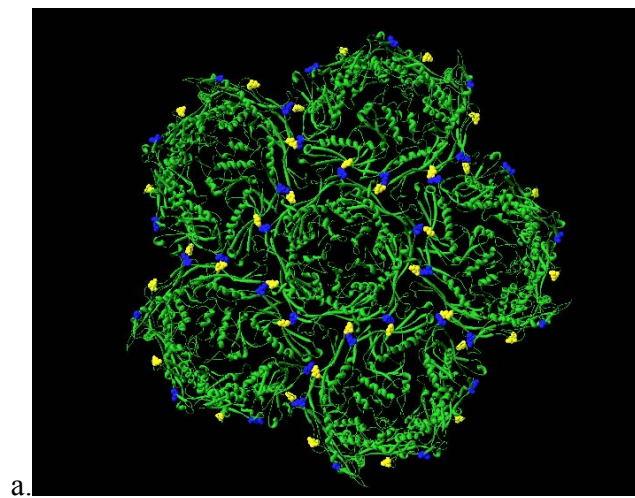
hexamers (Figure 34g, 34h, and Appendix D). The side-chain distance between the residues 231 and 178 from two capsomers on Prohead II is  $\sim 2\text{\AA}$  (Figure 34). This indicates that this unique interaction provides an ability to assemble a large structure *i.e.* proheads. These findings also support the previous results that the single mutants D231L and D231A cannot assemble proheads, since the particular amino acid, Asp (D), was not presented at residue 231 in the mutants D231L and D231A, which is able to pair with the Lys (K) at residue K178.

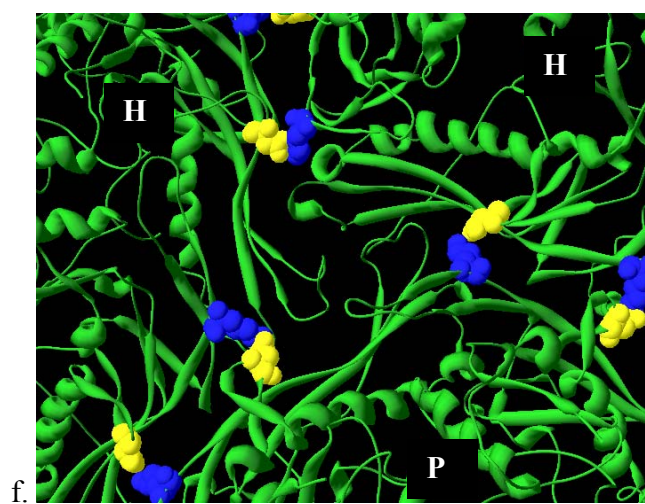
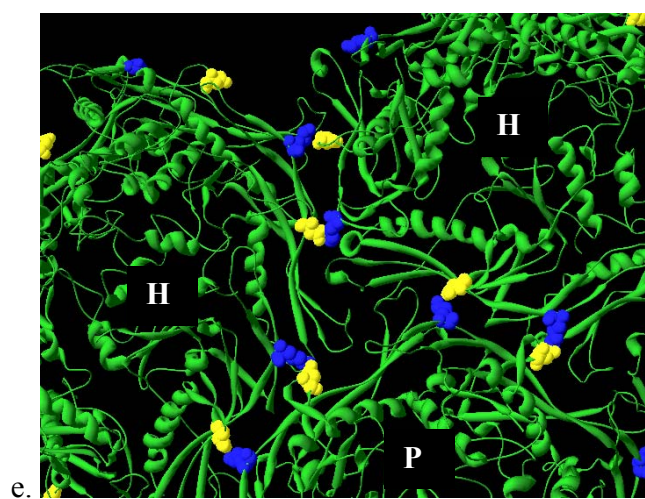
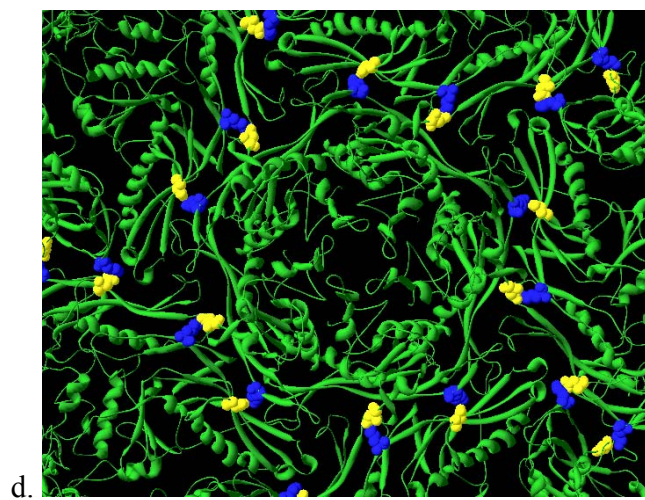


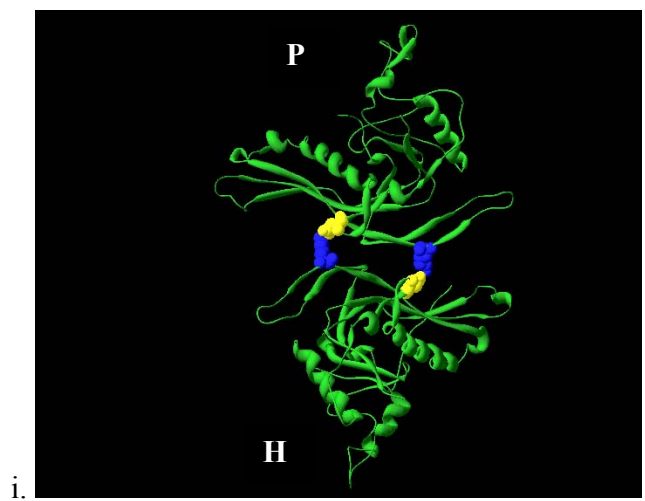
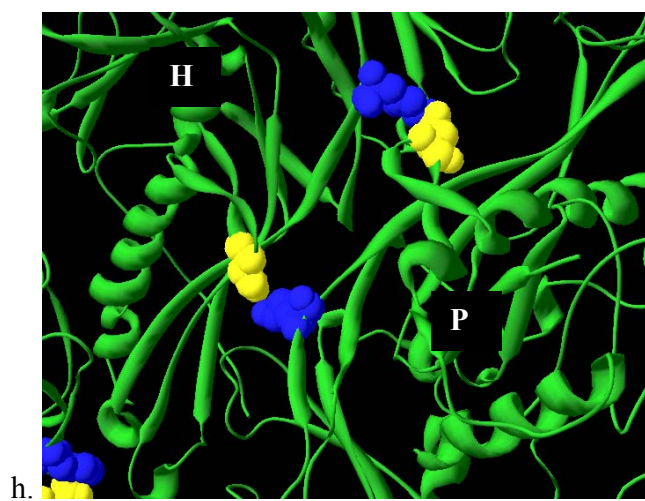
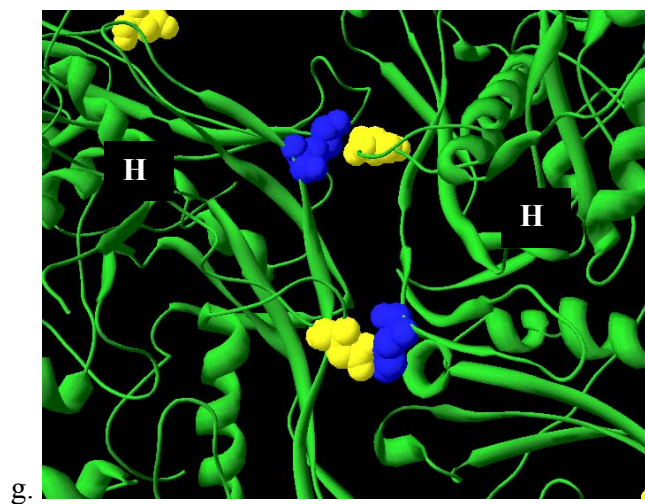
**Figure 33: The single subunit gp5 from the HK97 Prohead II (a) and the HK97 Head II (b) structures**

a. A folded single gp5 (green ribbon) from the HK97 Prohead II (PDB access number: 3E8K). b. A folded single gp5 (green ribbon) from the HK97 Head II (PDB access number: 1OHG ). The residues D231 and K178 are labeled in color yellow and blue.

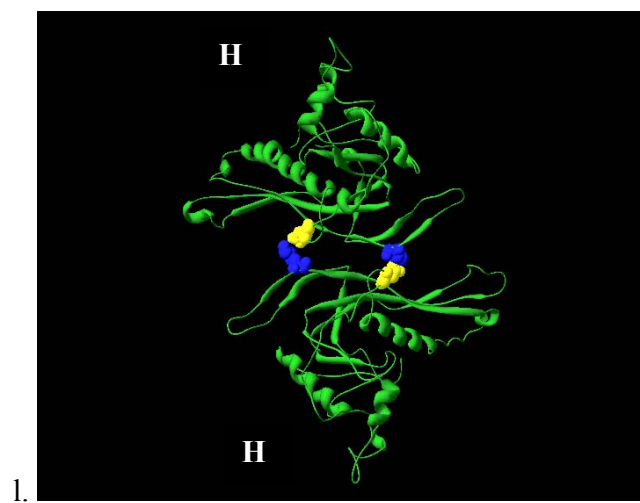
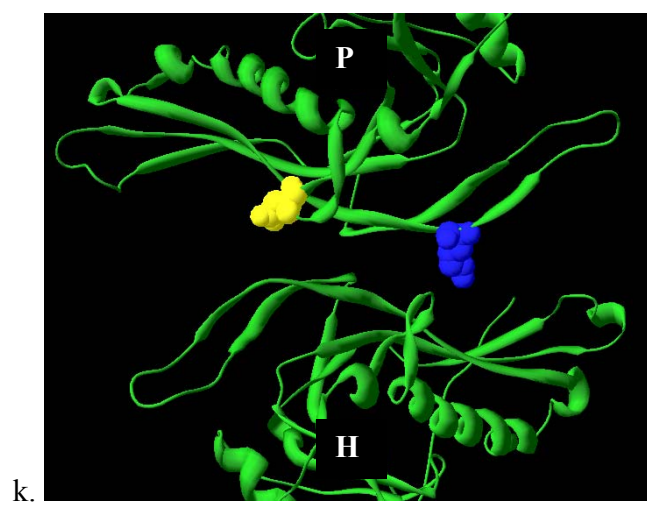
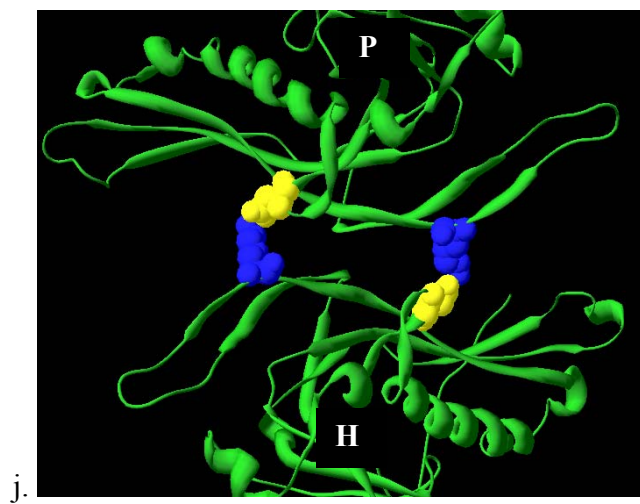


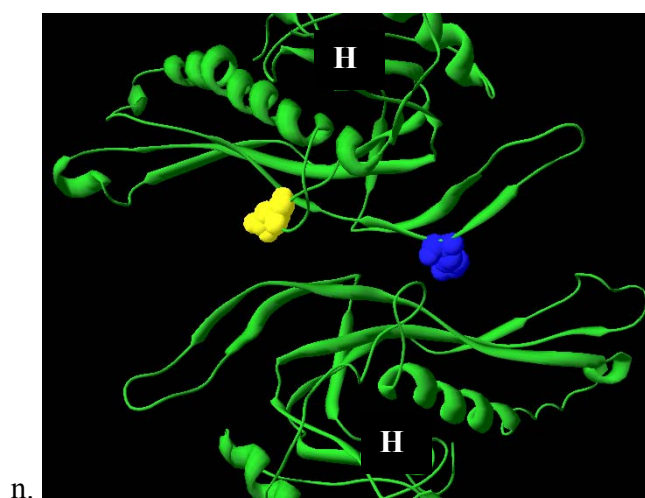
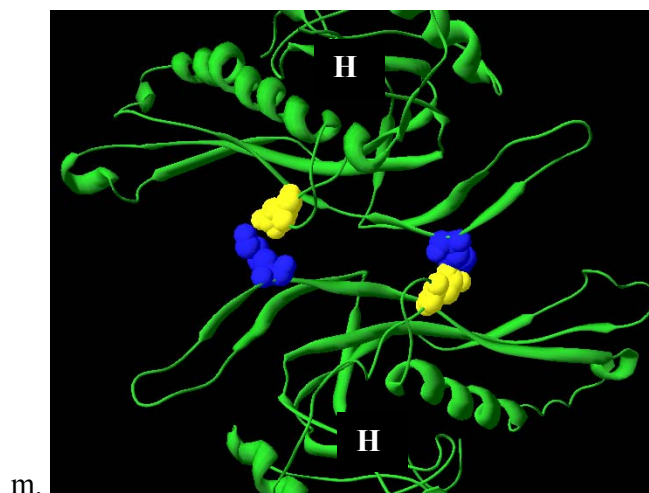












**Figure 34: The interactions of residues D231 and K178 between HK97 Prohead II capsomers**

a-h. The gp5 residue D231 interacts with the residue K178 on HK97 Prohead II (PDB access number: 3E8K). All of the residues D231 and K178 face out of each capsomer showing on one pentamer and five hexamers. b-h. The close-up views of the centered pentamer and surrounded hexamers. b and d. the close-up views of the centered pentamer. c. A close-up view of two hexamers. e and f. The close-up views of one hexamer-pentamer-hexamer junction. g. The residues D231 and K178 between two nearby hexamers. h. The residues D231 and K178 between a hexamers and a pentamer. i-m. The residues D231 and K178 are located on two single gp5's. i-k. The residues D231 and K178 are located in two hexamer single gp5's. l-n. The residues D231 and K178 are located in two single gp5's from a pentamer and a hexamers. m and k. One pair of residues D231 and K178 is shown. The HK97 gp5 is shown in green ribbon, and the residues D231 and K178 are shown in color yellow and blue.

### **The Prohead I D231L/K178N, D231L/K178I, and D231L/K178V cannot be purified and detected**

The relationship of residues 231 and 178 provides an essential interaction to “zip” the capsomers together. The ionic interaction of residue D231 and K178 has been seen in wild type HK97, while the hydrophobic interactions were found in two kinds of revertants: D231L/K178V and D231L/K178I. Genetic data showed that the hydrophobic interactions Leu-Val (L-V) and Leu-Ile (L-I) can replace the ionic interaction Asp-Lys (D-K) between residues 231 and 178 (Table 8), indicating that the characters of the gp5 of three revertants may be different from the wild type. To further study the three revertants, three double mutant gene 5's containing the double mutations D231L/K178N, D231L/K178I, and D231L/K178V were constructed in the plasmid pV0 for more intensive investigation.

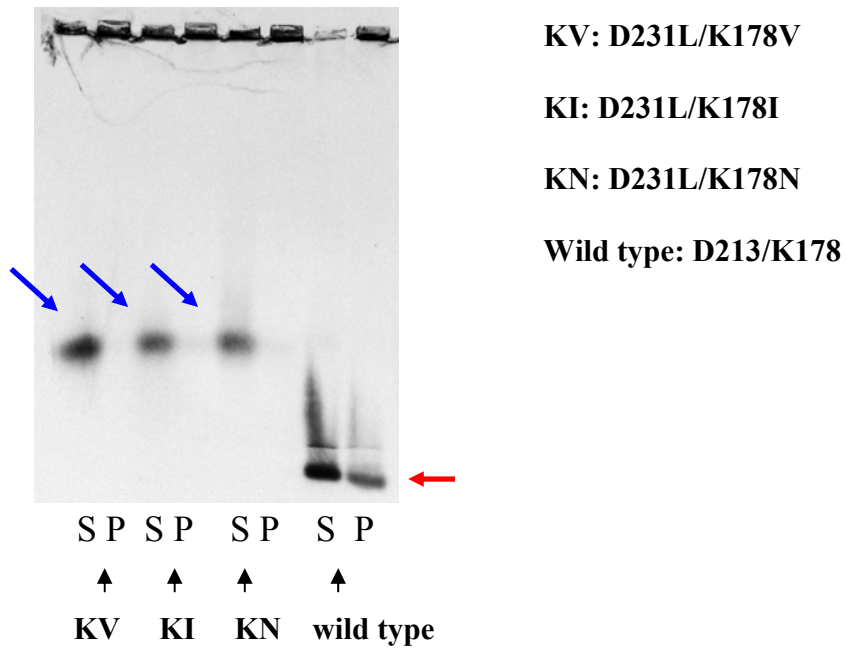
Previous data showed that HK97 Prohead I can be dissociated into capsomers after Prohead I was incubated with 2M KCl (Xie and Hendrix, 1995). I postulate that high salt favors Prohead I dissociation because it may break the salt bridge between capsomers in the prohead. Since two of the revertants have hydrophobic residues, L-V and L-I at the gp5 positions 231 and 178, the stability of these two proheads in high salt may be different from that of wild type.

To perform the dissociation assay, I attempted to purify Prohead I D231L/K178I. The same treatment on the two Prohead I's, D231L/K178V and D231L/K178I, may cause the same result since the relationship of residue 231 and 178 in the revertants D231L/K178V and D231L/K178I is hydrophobic interactions. I expected that the dissociation results of two Prohead I's, D231L/K178V and D231L/K178I, would be similar.

Unexpectedly, Prohead I D231L/K178I cannot be purified because of the failure of the salt and PEG precipitation step, suggesting that the three double mutant proheads are not stable

under our laboratory purification conditions. This indicates that double mutant gp5 D231L/K178I has different protein property than the wild type, either it cannot assemble Prohead I or its proheads are not stable enough to be purified.

To further examine the cause of the above phenomenon, three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N were expressed alone and radio-labeled with <sup>35</sup>S-Met. The collected protein samples were analyzed on an agarose gel to detect prohead formation (Figure 35). The results show that the three double mutant gp5's can self-assemble soluble capsomers, but no prohead can be detected from any of the double mutants (Figure 35). On the other hand, the wild type gp5 can assemble Prohead I under the same condition, and most of the proheads are in the soluble form. All of the above data illustrated that the three double mutants Prohead I are not stable, since they cannot be detected and purified.



**Figure 35: The double mutant Prohead Is D231L/K178N, D231L/K178I, and D231L/K178V are not detected**

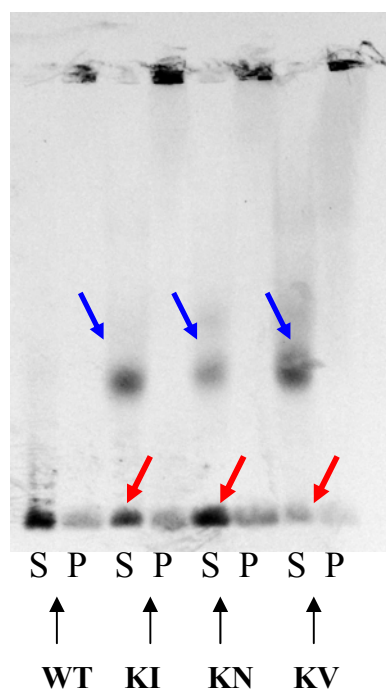
The double mutant gp5's D231L/K178N, D231L/K178I, and D231L/K178V, and the wild type gp5 were expressed alone and radio-labeled with  $^{35}\text{S}$ -Met. All of the mutant gp5's assemble soluble capsomers while the wild type gp5 assembles proheads. All protein samples were collected as the supernatant (S) and the pellet (P) and were analyzed on a 0.8% agarose gel. The capsomer and proheads are pointed with blue and red arrows.



### **Prohead II D231L/K178N, D231L/K178I, and D231L/K178V can be detected**

The HK97 wild type Prohead I can be dissociated into capsomers by 2M KCl, but the wild type Prohead II cannot be dissociated under the same conditions (Xie and Hendrix, 1995), indicating that the stability of the prohead has changed after the proteolysis. I hypothesize that Prohead II D231L/K178V, D231L/K178I, and D231L/K178N are more stable than their Prohead I.

If the above hypothesis is correct, three of the double mutant Prohead II may be detectable. To test the above assumption, three individual double mutant gp5's were radio-labeled with <sup>35</sup>S-Met and co-expressed with the gp4 for producing Prohead II. The results show that three double mutant proheads can be detected on an agarose gel, and the protease cleavage may provide prohead stability in the three revertants (Figure 36).



**KV: D231L/K178V**

**KI: D231L/K178I**

**KN: D231L/K178N**

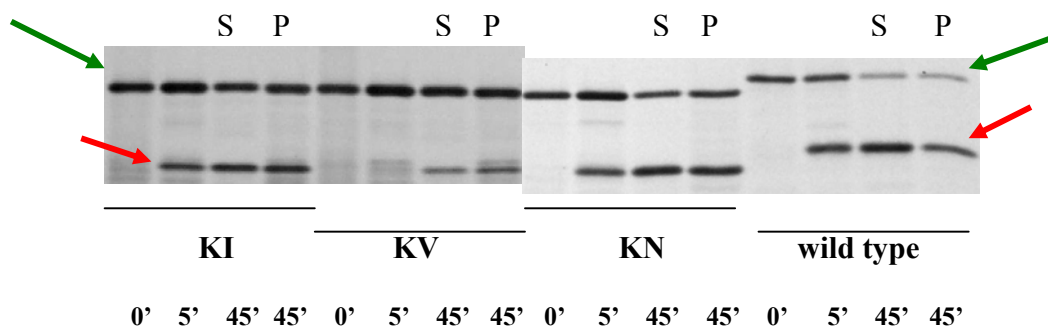
**Wild type: D213/K178**

**Figure 36: The double mutant Prohead IIs D231L/K178N, D231L/K178I, and D231L/K178V are detectable**

The double mutant gp5's D231L/K178N, D231L/K178I, and D231L/K178V, and the wild type gp5 were co-expressed with the gp4 alone and radio-labeled with  $^{35}\text{S}$ -Met. All samples were collected as the supernatant (S) and the pellet (P), and were analyzed on a 0.8% agarose gel. The capsomers and proheads were pointed by the blue and red arrows.

The wild type Prohead II is stable under the laboratory conditions and the wild type prohead assembly and the proteolysis occur rapidly. There is no observed dissociated wild type capsomer or unassembled capsomer (Figure 36). Interestingly, some of the mutant capsomers can be detected, indicating that the properties of the prohead D231L/K178V, D231L/K178I, and D231L/K178N are different from the wild type and these differences may influence capsid assembly (Figure 36).

To determine whether the gp4 can cleave these three mutant gp5's and also to identify the proheads shown the agarose gel as the double mutant Prohead II, three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N were analyzed on an SDS PAGE to monitor their proteolytic cleavage. The results show that these three gp5's can be partially cleaved by the gp4, confirming that the proheads which were detected on the agarose gel should be Prohead II (Figure 37).



**KV: D231L/K178V**

**KI: D231L/K178I**

**KN: D231L/K178N**

**Wild type: D213/K178**

**Figure 37: Proteolysis of three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N**

The double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N, and the wild type gp5 were co-expressed with the gp4 alone and radio-labeled with  $^{35}\text{S}$ -Met. All samples were collected after incubating for 0min, 5min, and 45min and were analyzed on a 10% SDS PAGE. Three double mutants can be cleaved by the gp4. The supernatant (S) and pellet (P) proteins were collected after 45min. The cleaved and un-cleaved gp5 are pointed by the red and green arrows.

Interestingly, three double mutant gp5's produce smaller amounts of Prohead II than the wild type (Figure 37), suggesting that three double mutants may assemble their Prohead II less efficiently than the wild type, possibly due to the different interactions of the residues 231 and 178 between the mutant capsomers. Furthermore, lower amounts of the cleaved double mutant gp5's were observed on the SDS PAGE (Figure 37), also confirming that the protease cleavage, as well as Prohead II assembly, is less efficient in the revertants than the wild type.

Another interesting finding is that the three gp5 double mutants D231L/K178V, D231L/K178I, and D231L/K178N have different levels of deficiency in assembly of Prohead II, shown on the agarose gel in Figure 37. The apparent efficiency of Prohead II assembly decreases in the order D231/K178 (wild type) > D231L/K178N > D231L/K178I > D231L/K178V (Figure 37).

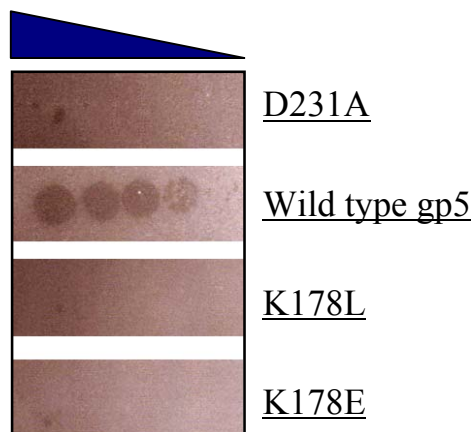
According to the proteolysis monitored on the SDS PAGE, the amounts of cleaved gp5 from double mutant gp5 D231L/K178N is more than from D231L/K178I, and D231L/K178V performs the least protease cleavage (Figure 37). These findings are consistent with the different amounts of the mutant Prohead II production detected on the agarose gel (Figure 36). The efficiencies of the cleavage are as follows: D231/K178 (wild type) > D231L/K178N > D231L/K178I > D231L/K178V (Figure 37). Nevertheless, since the un-cleaved gp5 can be detected on the SDS PAGE (Figure 37), the capsomers detected on the agarose gel are un-assembled capsomers, not dissociated capsomers from Prohead II.

All of the results show that three double mutants and wild type gp5 are functional in the viable phages, but have differences in the early stages of the HK97 capsid assembly.

### **The single mutant gp5's K178E and K178L cannot produce functional gp5**

To test whether the single mutants K178E and K178L can produce functional gp5, two single gp5 mutants K178E and K178L were constructed in the plasmid pV0 for determining the biological activity of their gp5's. The results show that neither mutant can complement gene 5<sup>-</sup> amber phage, indicating that neither K178E nor K178L produce functional gp5 (Figure 38).

### **The gene 5<sup>-</sup> amber phage**



**Figure 38: The single gp5 mutants K178E and K178L do not produce functional gp5**

The single mutants K178L, K178E, D231A, and the wild type gp5 are carried in the plasmid pV0 alone and expressed in the expression strain BL21(DE3). Serially diluted gene 5<sup>-</sup> amber phage was spotted on the above bacterial lawns. The results from the single mutant D108A and the wild type gp5 are shown as the negative and the positive controls.

### **The gp5 single mutants D231L, D231A, K178E, and K178L do not assemble proheads**

To test whether the single mutants D231L, D231A, K178E and K178L can assemble capsomers and proheads, these mutants and the wild type gp5's were labeled with  $^{35}\text{S}$ -Methionine and expressed alone in *E. coli*. The proteins were collected and examined on an agarose gel. The results show that all four single mutants can assemble capsomers but not proheads, indicating that residues D231 and K178 are both necessary in the wild type HK97 to assemble capsomers into proheads (Figure 30a). Furthermore, the mutant capsomers D231L, D231A, K178E and K178L show different surface charges, suggesting that residues 231 and 178 face out of the capsomers. The capsomer production of these mutants was confirmed by examining all of the samples on a native gel. The pentamers and hexamers of the single mutants D231L, D231A, K178E and K178L are monitored on a native gel. All of the mutant capsomers have different surface charges than the wild type (Figure 30b).

**The gp5 single mutant proteins D231L, K178E, and K178L cannot be cleaved by the protease gp4**

To assay for proteolysis of the gp5 mutants D231L, K178L and K178E, the protein samples were collected and examined on an SDS PAGE. The results show that no protease cleavage can be detected from these mutant gp5's (Figure 37), suggesting that the single mutations D231L, K178E, and K178L in gp5 can block the protease cleavage.

An unexpected mobility shift was observed when examining the mutant gp5 D231L and D231A on an SDS PAGE (Figure 32 and 37). Both denatured gp5's D231L and D231A migrate faster than wild type gp5 (D231) on a 10% SDS PAGE but the migration level of these two mutants are very similar to each other. The double mutant gp5 D231L/K178I has a different mobility shift than wild type gp5 as well, indicating that mutant D231L/K178I retains certain properties of the single mutant gp5 D231L (data not shown). This alteration of the residue 231 is a residue-specific phenomenon since this shift was not observed on other gp5 mutants, e.g. K178L and K178E (Figure 38). A similar altered mobility has been detected from the P22 spike protein mutants e.g. H304 and tsfG244R (Fane et al., 1991; Villafane, Fleming, and Haase-Pettingell, 1994). Interestingly, the suppressor mutants RH304.8, RH304.14, and RH304.18, retain the altered migration (Villafane, Fleming, and Haase-Pettingell, 1994). The reason of these SDS PAGE mobility shift were unknown.



## 5.4 SUMMARY & DISCUSSION

### **The essential pair of the residues 231 and 178 in the HK97 gp5**

At the beginning of my research, I have constructed several different mutant lysogens for isolating revertants. I planned to determine the range of variation in the recombination efficiency and the reversion frequency, but most importantly to gain access to a variety of functional aspects of the capsid protein. I have selected some gp5 residues based on their unusual location on the HK97 capsid and their predicted functional roles. An example of such a residue is D231, described in this chapter. The gp5 T loop is a unique structure and its functions were unknown. The outermost residue 231 in the T loop was selected for examining its biological function *in vivo*. Surprisingly, it was found that residue 231 plays a central role in the HK97 capsid assembly.

Based on earlier studies on isolating suppressors in the phage lambda and P22 coat proteins, I expected to isolate suppressors which are located in a special region of gene 5 as a functional domain. The suppressors in the lambda gene *E*, which encodes the coat protein, can rescue a phage lambda mutant with a defective gpFI (Murialdo and Tzamtzis, 1997). These intergenic suppressors were found in a special region of gene *E* which was named the EFi domain (Murialdo and Tzamtzis, 1997): this was taken as evidence to define the EFi domain as a specific region of gpE that interacts with gpFI. Another example is the suppressor mutants of phage P22. The single amino acid substitutions D163G, T166I, and F170L were isolated to globally suppress the folding defects of temperature-sensitive (ts) mutants in the phage P22 coat protein (Aramli and Teschke, 1999). These three intragenic suppressors are located close to each

other in the primary sequence, suggesting that the region around the residues 163 to 170 of the P22 coat protein may have important roles in proper function of the coat protein probably at the level of protein folding.

It was not surprising to see an intragenic suppressor complement a lethal mutation in gp5, but it is interesting to find that all of the isolated second-site suppressors map to the same position, i.e. residue 178 (Table 8). Some particular domains of gp5, *e.g.* the E loop, are responsible for directing specific aspects of capsid assembly and maturation (Lee et al., 2008). However, I did not discover a functional region of HK97 gp5. Instead of identifying a particular domain in the HK97 gp5, single substitutions at residue 178, *e.g.* K178V, K178I, and K178N, were found to suppress the single substitution at residue 231, D231L. The pair of the gp5 residues 231 and 178 on the HK97 capsid was defined as a crucial interaction for assembling capsomers into proheads in my studies.

Some suppressors were found to only partially rescue a lethal mutation in an essential gene, *e.g.* the suppressor P65Q of the single substitution R170Q in the Rous Sarcoma Virus (RSV) capsid protein (CA) (Bowzard, Wills, and Craven, 2001). But the second-site suppressors K178V, K178I, and K178N can fully suppress the single substitution D231L in the HK97 gp5 without making a big influence in the original bacteriophage HK97 capsid assembly (see below).

### **Residues 231 and 178 play a crucial role in the HK97 capsomers' functionality**

The above study showed that non-functional gp5 was produced when the gp5 carries a single substitution, either D231L or D231A (Figure 31 and 38), implying that residue D231 is essential for producing functional gp5 in the wild type HK97. These single mutant gp5's assemble capsomers but not proheads (Figure 30), indicating that the above single substitutions may block the functional roles of residue 231 in the step of assembling capsomers into proheads. Furthermore, the mutant capsomers D231L and D231A migrate slower than the wild type capsomer shown on the agarose gel and the native gel (Figure 30), possibly due to the loss of the negative charges on the mutant capsomers. These two mutant capsomers D231L and D231A have different surface charges than the wild type capsomer since residue 231 faces out of the capsomer to possibly alter the surface charge of the capsomer.

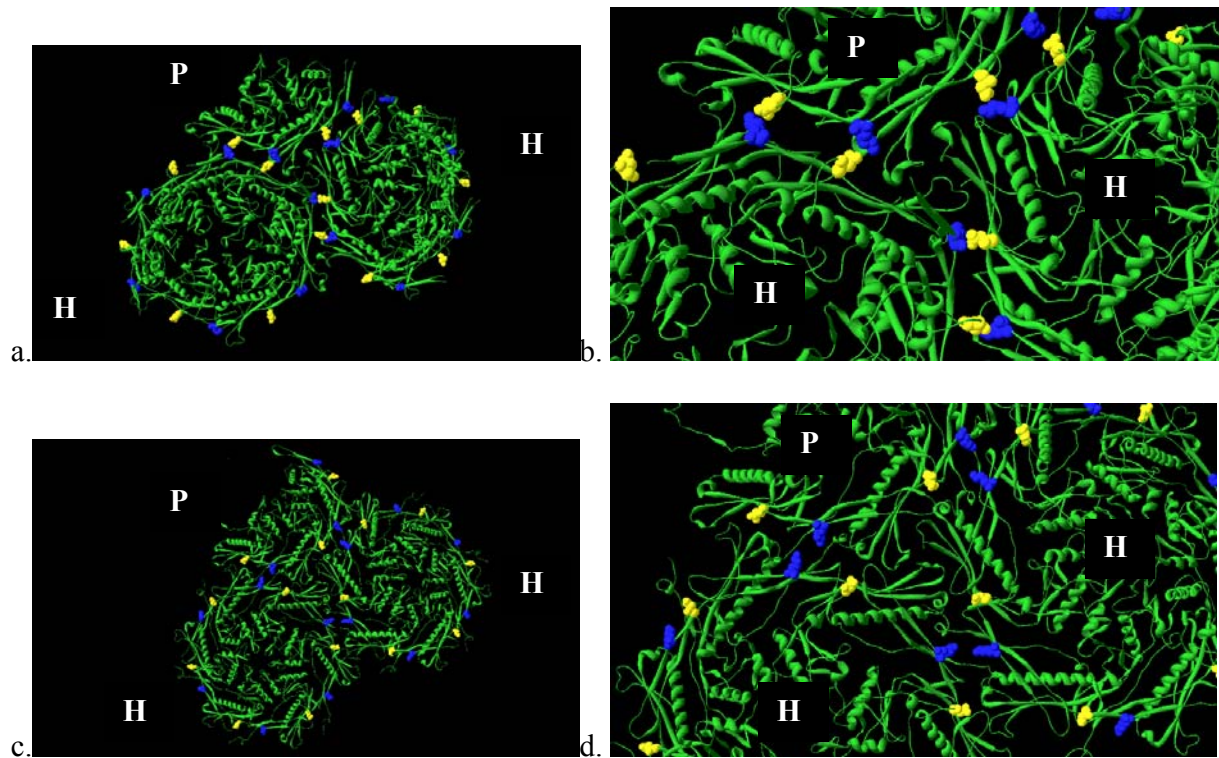
The residue 178 was studied after the relationship of residue 231 and 178 was determined. The complementation results showed that two single mutants K178L and K178E cannot produce functional gp5 (Figure 38), indicating that the residue K178 in the wild type gp5 is essential. This is different from some of the isolated suppressors. For examples, some of the suppressor mutants alone in the phage P22 (Parent, Suhanovsky, and Teschke, 2007) and the virus RSV (Lokhandwala et al., 2008) do not have any deleterious phenotype. Furthermore, the mutant gp5's K178L and K178E can assemble capsomers but not proheads (Figure 30a), suggesting that a single substitution at residue 178 is able to block the pathway of assembling capsomers into proheads. These results demonstrate that residue 178 is involved in the pathway of assembling capsomers into proheads.

The surface charges of the mutant capsomers K178L and K178E are different from the wild type capsomer. But unlike the above D231L and D231A, the mutant capsomers K178L and

K178E migrate faster than the wild type: K178E > K178L > K178(=wild type) (Figure 35). This may be the loss of the positive charge (K178L) and the gain of the negative charge (K178E) on the mutant capsomers. These data show that the charge of the residue 178 affects the surface charge of the assembled capsomers, and confirm that residue 178 is not buried inside but exposed from capsomers. Thus, the surface charges of the capsomers can be influenced by the charges of the residues 231 and 178.

**The residue 231-178 interaction is not required after HK97 capsid assembly and maturation**

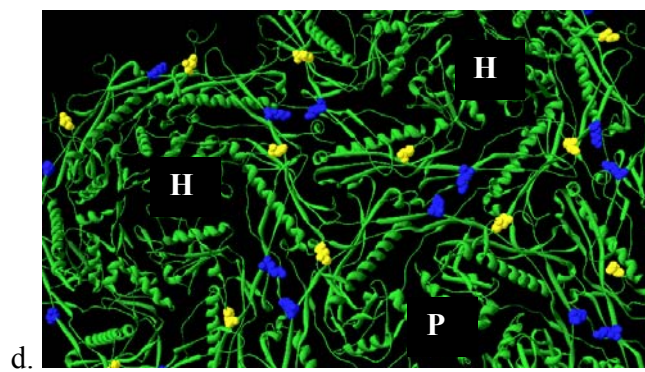
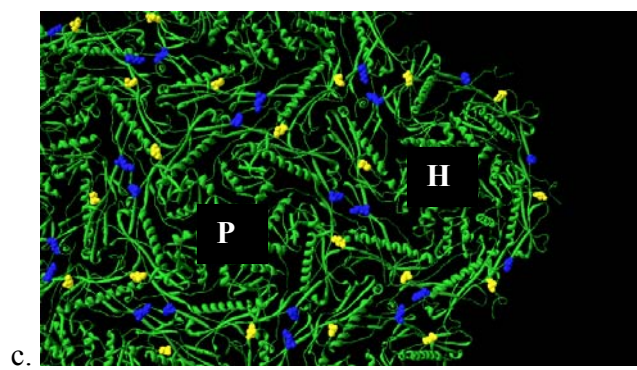
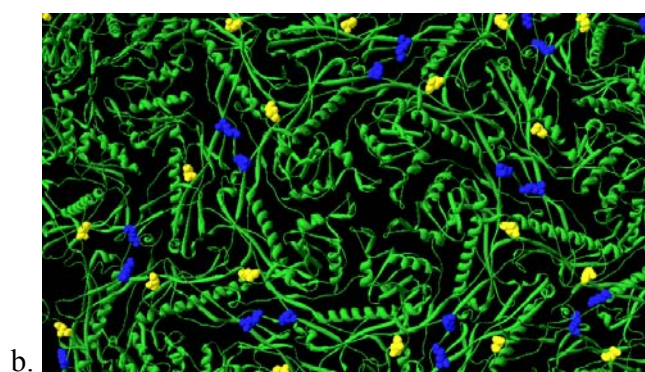
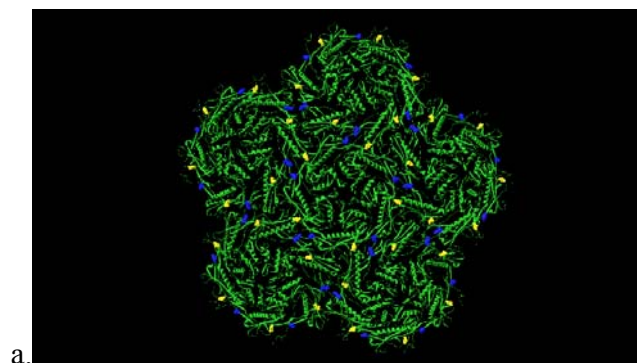
The above data confirmed the model for the interaction of residues 231 and 178 on the HK97 Prohead II. The residues D231 and K178 were found to locate at the margin of the capsomer (Figure 34, 39, Appendix D) and such interaction was between two gp5's located in two capsomers, not between two gp5's in the same capsomers (Figure 34, 39, Appendix D). It can be found either between two hexamers, or between pentamers and hexamers (Figure 34, 39, Appendix D). This interaction may not be possibly applied between the residue D231 and K178 within a subunit gp5 ( $\sim 26\text{\AA}$ ) (Figure 33).



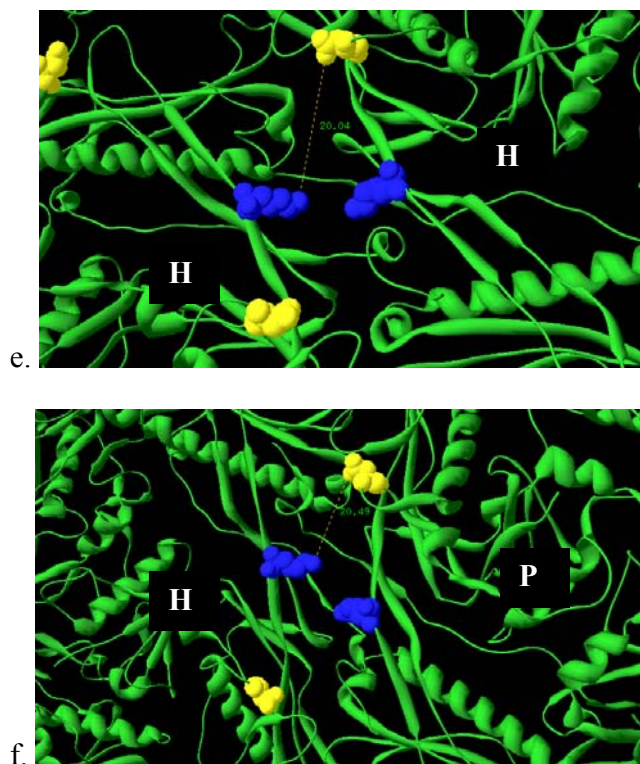
**Figure 39: The residues D231 and K178 on the HK97 Prohead II (a and b) and the HK97 Head II (c and d) structures**

a. and b. The residues D231 and K178 on the HK97 Prohead II. c. and d. The residues D231 and K178 on the HK97 Head II. The green ribbon is the HK97 subunit gp5. The residue D231 is labeled in color yellow and the residue K178 is labeled in color blue.

There is no high resolution HK97 Prohead I structure currently available, so the crystal structural of Prohead II was used to analyze the interaction of residues D231 and K178 (Gertsman et al., 2009). Based on the existing cryo-EM structures, there is not much structural difference between Prohead I and Prohead II except for the Delta domain (inside the Prohead I shell) (Conway et al., 1995). The prediction from my genetic experiments that D231 and K178 make a salt bridge between capsomers is supported by the recently determined X-ray structure of Prohead II (Gertsman et al., 2009). Interestingly, this structural relationship between residue D231 and K178 is not found in the HK97 Head II (Figure 39 and 40). After conformational changes, the two residues D231 and K178 were translocated far apart from each other (Figure 39 and 40). The distances between the side chains of residues D231 and K178 in two adjacent capsomers is  $\sim 20\text{\AA}$  (Figure 39 and 40). This was very different from the corresponding distances found in the HK97 Prohead II structure ( $< 2\text{\AA}$ ) (Figure 34). This implies that while the interaction of residues 231 and 178 between the capsomers may be necessary for HK97 capsid assembly, the HK97 capsid apparently does not require these interactions after capsid assembly and maturation are complete.







**Figure 40: The relationship of the residues D231 and K178 on the HK97 Head II**

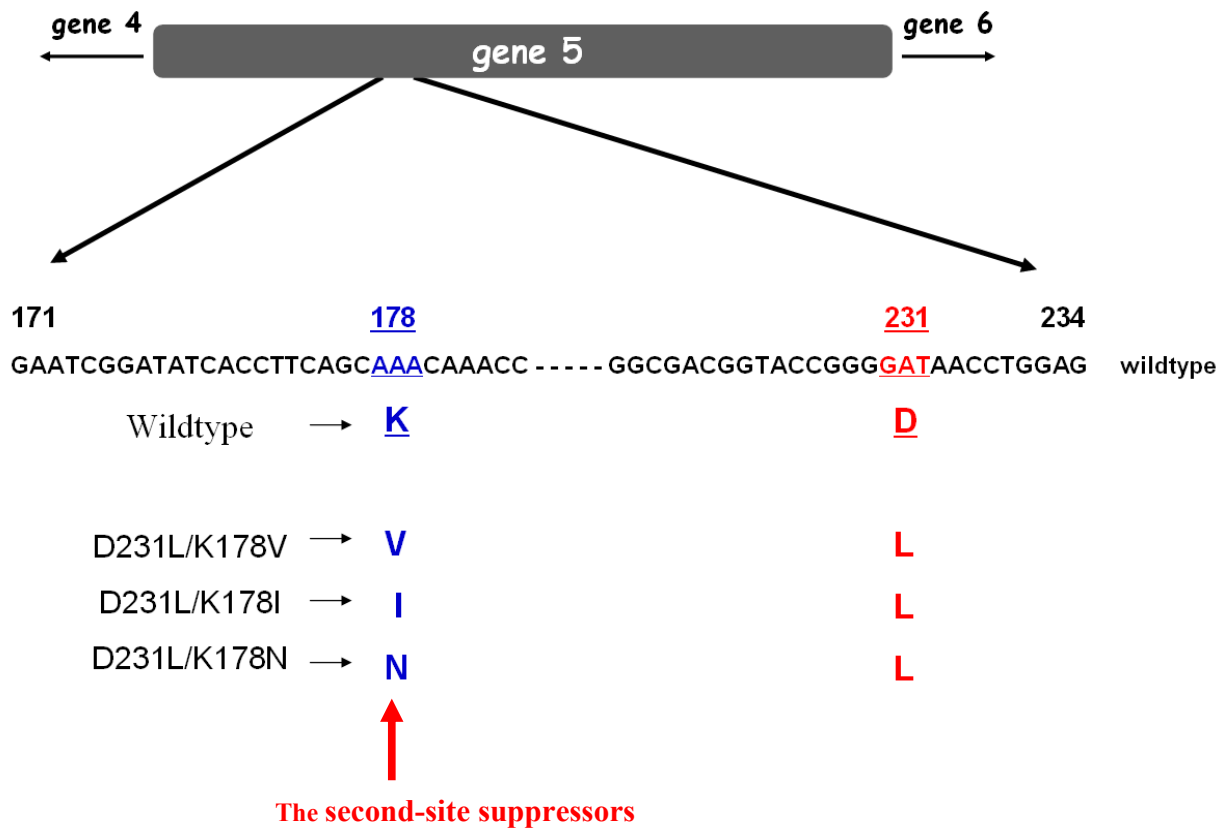
a-f. The location of the residues D231 and K178 on the HK97 Head II (PDB access number: 1OHG). a. All of the residues D231 and K178 face out from each capsomer showing on one pentamer and five hexamers. b. A close-up view of one pentamer. c. The relationship of the residues D231 and K178 between a pentamer and a hexamers. d. A close-up view of the hexamer-pentamer-hexamer junction. e. The distance of two residues, D231 and K178, between two hexamers ( $\sim 20\text{\AA}$ ). f. The distance of two residues, D231 and K178, between a pentamer and a hexamer ( $\sim 20\text{\AA}$ ). The HK97 gp5 is shown in green ribbon, and the residues D231 and K178 are shown in color yellow and blue.

### **The alternative amino acid pairs at positions 231 and 178**

The results reported here argue that wild type HK97 has a pair of residues D231 and K178 that form a salt bridge (an ionic interaction) between the capsomers which promotes assembling capsomers into proheads. Genetic results showed that three alternative pairs of residues 231 and 178, D231L/K178V, D231L/K178I, and D231L/K178N, can support production of viable bacteriophage HK97 (Table 8), indicating that they are able to provide the functions in assembling capsomers into proheads.

Three second-site suppressors, K178V, K178I, and K178N, were isolated from multiple experiments (Table 8). Two of the three suppressors, K178V and K178I, are different from the other suppressor K178N in the revertants (Table 8). The D231L/K178V (L-V) and D231L/K178I (L-I) interactions are hydrophobic, but the D231L/K178N (L-N) interaction is not (Table 8). The revertants D231L/K178N has been isolated from two individual experiments with two different codons, N(AAT) and N(AAC), and the complementation results confirmed that Asn (N) at residue 178 (K178N) can suppress Leu (L) at residue 231 (D231L). These data strongly support the Asn suppression of Leu in gp5. The phenotypes of the revertant D231L/K178N are not different from other two revertants D231L/K178V and D231L/K178I. I conclude that K178V, K178I, and K178N are the major suppressors of the single substitution D231L in the HK97 gp5.

The HK97 gp5 which fits the theory of quasi-equivalence can be plastic enough undergo several different conformational changes. This implies that bacteriophage HK97 can bear certain substitutions in the gp5 sequence, *e.g.* an alternative pair of residue 231 and 178. Thus, it is possible that two single nucleotide substitutions in the revertants D231L/K178V, D231L/K178I, and D231L/K178N may produce slightly different HK97 capsid than the wild type, but the HK97 gp5 can bear these differences.



**Figure 41: The sequence pairs of the gp5 residues 231 and 178 in the wild type HK97 and the three revertants**

The partial nucleotide sequence of HK97 gp5 is listed from residue 171 to 234.

### **Phenotypes of double mutants D231L/K178V, D231L/K178I, D231L/K178N**

The previous data showed that wild type gp5 and three double mutants produce different amounts of proheads and have differences in the extent of proteolysis (Figure 36 and 37), indicating that there is a difference among the wild type HK97 and three revertants related to the pathway of assembling capsomers into proheads. It was found that three double mutants, that provide the residue 231-178 interactions L-V, L-I, and L-N in the revertants, do not produce Prohead I that can be detected *in vitro* (Figure 35) when the three double mutant gp5's D231L/K178V D231L/K178I, and D231L/K178N were expressed alone in *E. coli*. Several different purification conditions have been used to attempt to purify Prohead I from one of the double mutants, D231L/K178I (data not shown), but there was no visualized and purified Prohead I D231L/K178I *in vitro*. However, the fact that these double mutants produce Prohead II when protease is present strongly argues that Prohead I must have been produced in *E. coli* and been stable long enough to be converted to Prohead II. The failure to find Prohead I *in vitro* in the absence of protease therefore most likely reflects instability of the double mutant Prohead I in the *in vitro* buffers used.

These three revertants are not temperature sensitive mutants (data not shown), and there was no significant growth defect when these revertants were grown on the *E. coli* lawn Ymel (data not shown). Besides, the phage titers of these three revertants do not decrease after a serial passage (data not shown). These results suggest that three revertants do not accumulate significant phenotypic defects. At least it cannot be observed in my research. Furthermore, to determine whether the three revertants have any defect in the late capsid maturation stages, Prohead II from mutant D231L/K178I was purified for further examination. There was no observed defect when the Prohead II D231L/K178I was examined for expansion and

crosslinking *in vitro* (data not shown), suggesting that three revertants D231L/K178V, D231L/K178I, or D231L/K178N may produce viable bacteriophage HK97 particles without accumulating defects during maturation.

Beyond prohead assembly, none of the three revertants D231L/K178V, D231L/K178I, and D231L/K178N have distinguishable phenotypes differing from the wild type. Thus, all of the data suggest that the difference among the wild type and three revertants D231L/K178V, D231L/K178I, or D231L/K178N may relate only to the pathway of assembling capsomers into proheads. I conclude that there is a slight effect on the capsid assembly when the residues D-K in the wild type gp5 were substituted with the pairs of L-V, L-I, or L-N in the gp5, but the alternative pairs of residue 231 and 178 do not cause a big alteration.

### **Proteolysis stabilizes the HK97 proheads in the revertants, D231L/K178V, D231L/K178I, and D231L/K178N**

In my study, I discovered that three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N alone in *E. coli* cannot assemble proheads, but the wild type gp5 expressed alone does assemble proheads (Prohead I) (Figure 35), suggesting that the assembled precursor capsids (Prohead I) of the three double mutants were not stable in our laboratory conditions, apparently less stable than the wild type proheads, and may need additional aid for stabilizing the precursor capsid. However, three double mutant gp5's D231L/K178V, D231L/K178I, and D231L/K178N and the wild type gp5 can assemble Prohead II when co-expressed with gp4. This suggests that three double mutant Prohead II's are more stable than double mutant Prohead I and proteolysis may contribute to the prohead stability in three HK97 revertants D231L/K178V, D231L/K178I, and D231L/K178N.

No un-assembled capsomer or dissociated capsomer was detected when the wild type gp5 was expressed alone or expressed with the protease gp4. Prohead I and Prohead II were detected on the agarose gel, respectively (Figure 35 and 36), indicating that wild type gp5 is able to fully assemble Prohead I and Prohead II and these proheads are stable. However, the Prohead II D231L/K178V, D231L/K178I, and D231L/K178N can be assembled but there were some capsomers detected on the agarose gel (Figure 36). These capsomers were confirmed to be un-cleaved capsomers since the un-cleaved gp5's were detected on the SDS PAGE (Figure 37). The results show that the order of the amounts of the Prohead IIs shown on the agarose gel were the same as the order of the cleaved gp5's shown on the SDS PAGE: D231L/K178V < D231L/K178I < D231L/K178N < D231L/K178 (wild type) (Figure 36 and 37). The dissociated capsomers from the Prohead II should be all shown as the cleaved gp5's on SDS PAGE.

However, there are the cleaved and un-cleaved gp5's of double mutants. It was known that three double mutants cannot assemble Prohead I. Thus, the cleaved gp5's shown on SDS PAGE are from the assembled Prohead II and the un-cleaved gp5's are the un-assembled capsomers. This suggests that the proteolysis in three revertants provides the ability to hold the prohead structure together. The data also suggest that three double mutant Prohead II's are sufficiently strong enough since the capsomers do not dissociate from the Prohead II's D231L/K178V, D231L/K178I, and D231L/K178N.

All of the data indicate that the non-covalent L-V, L-I, and L-A interactions of the gp5 residues 231 and 178 in the revertants provide the prohead assembly less stability than the wild type Asp-Lys interaction. Thus, the ionic interaction of the residues 231 and 178 may be more efficient than other non-covalent interactions, *e.g.* hydrophobic interaction. Besides, the proteolysis (the gp4 protease cleavage) is important to stabilize the residue 231-178 interactions in the revertants D231L/K178V, D231L/K178I, and D231L/K178N.

### **The functions of the gp5 T loop and residue 231**

The gp5 T loop contains nine residues, so it is not surprising that some of, or all of, the 9 residues—GDGTGDNLE—are essential (Figure 28 and 40, Appendix B). The previous data showed that mutant  $\Delta$ loop can produce gp5's, but cannot assemble these gp5's into capsomers (Figure 29 and 30), indicating that the T loop deletion blocks the HK97 capsid assembly pathway at a very early stage. The complementation results showed that mutant  $\Delta$ loop produces no functional gp5 (data not shown), suggesting that the T loop is essential for gp5 functionality and it is necessary for HK97 capsid assembly.

The single mutant gp5's, D231A and D231L, contain the entire T loop but carry a single substitution either Ala or Leu at residue 231. These two single mutants can assemble capsomers (Figure 30), indicating that the residues in the T loop are required for assembling the subunit gp5's into capsomers, but residue 231 may not be important for assembling the gp5's into capsomers. The capsomer assembly requires T loop but is independent of the residue 231 functionality. However, the mutant capsomers, *e.g.* D231L, and D231A, do not have the functions for assembling capsomers into proheads (Figure 30), indicating that they may not have the normal capsomers' functions although these mutant capsomers can be assembled. Thus, single substitutions D231L and D231A destroy the capsomers' functionality but do not block the capsomers formations.

Although a capsid protein sequence alignment suggests that some dsDNA bacteriophages probably have a loop region similar to the HK97 T loop. The amino acid sequence of the gp5 T loop was not found to be strongly conserved among dsDNA bacteriophages: HK97, D3, OP1, XP10, XOP411, BFK20, and Phi1026b (Figure 40). Some of the phages carry the charged residue Asp (D), but some of them do not (Figure 40). After examining the entire sequence of the



T loop, some of the other residues, i.e. 226, 228, and 233 in the T loop are fully conserved (Figure 40), suggesting that this domain region within these phages may have similar structures or functions.

Because the relationship between residues 231 and 178 was found in this study, I also examined the protein sequences around residue 178 in gp5 (Figure 40). Since the phages HK97, Xp10, Xop411, and OP1 carry an Asp (D) at residue 231, I expected to find that the corresponding residue 178 in these phages may have a similar amino acid like Lys (K) at residue 178 or around residue 178 in their major capsid protein. Interestingly, there is Lys(K) at the residue 187 in the phages Xp10, Xop411, and OP1, which are relatively aligned with the HK97 gp5 residue K178. This suggests a correlation between the relative residues 231 and 184 (the residue 184 for the phages other than HK97) in the dsDNA bacteriophages (Figure 40).

Although the phages BFK20, D3 and Phi1026b do not have the correlation of Asp-Lys (D-K) (salt bridge) similar to the phage HK97, the residues in these phages corresponding to HK97 residues 231 and 178 have another different correlation similar to the three isolated HK97 revertants. After further examining the aligned sequences of phages BFK20, D3 and Phi1026b, the correlation relationship were found to be non-ionic interactions: the Asn-Leu (N-L) in phage BFK20, the Ala-Leu (A-L) in phage D3, and the Ala-Ile (A-I) in phage Phi1026b. Future experiments on these phages are required to determine the functions of these residues.

The structure comparison among the dsDNA bacteriophages may facilitate understanding whether the dsDNA bacteriophages have a similar capsid assembly to HK97. However, none of the above dsDNA bacteriophages have solved structures of the mature capsid (or the precursor capsid) available in the protein database. Structural data for these phages would be helpful to

determine whether the interaction of HK97 gp5 residues 231 and 178 is a universal feature of the dsDNA bacteriophages' capsid assembly.

gene 4

gene 5

gene 6

176 177 178 179 180 226 227 228 - - 229 230 231 232 233 234

HK97

D3

OP1

XP10

XOP411

MU1/6

11B

BFK20

SFV

PHI1026B

<u>F</u>	<u>S</u>	<u>K</u>	<u>Q</u>	<u>T</u>	<u>G</u>	<u>D</u>	<u>G</u>	-	-	<u>T</u>	<u>G</u>	<u>D</u>	<u>N</u>	<u>L</u>	<u>E</u>
F	E	L	E	N	G	N	G	-	-	T	G	A	N	L	H
F	A	K	K	T	G	T	G	-	-	A	N	D	G	L	L
F	A	K	K	T	G	T	G	-	-	A	N	D	G	L	L
F	A	K	K	T	G	T	G	-	-	A	N	D	G	L	L
F	D	T	I	T	G	N	G	S	T	E	M	Q	G	I	L
I	S	M	I	D	-	-	-	-	-	-	-	-	-	-	A
F	D	I	V	T	G	D	G	-	-	T	G	N	N	L	T
F	G	M	G	S	G	T	G	A	G	T	P	K	Q	P	K
F	N	L	K	N	G	D	G	-	-	T	G	A	N	I	L

**Figure 42: The partial sequence alignment of the major capsid proteins from the dsDNA phages HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B**

The major capsid proteins of nine dsDNA bacteriophages have similarities to *E. coli* bacteriophage HK97 gp5 (385a.a.). The nine bacteriophages are *Pseudomonas* D3 (395a.a.), *Xanthomonas* phage OP1 (390a.a.), *Xanthomonas* phage Xp10 (390a.a.), *Xanthomonas* phage Xop411 (390a.a.), *Streptomyces* phage Mu1/6 (419a.a.), *Flavobacterium* phage 11b (379a.a.), *Corynebacterium* phage BFK20 (413a.a.), Enterobacteriophage SFV (409a.a.) and *Burkholderia* phage phi1026b (418a.a.). The sequences were analyzed by ClustalW (Appendix B). The residue numbers above the protein sequence correspond to the HK97 gp5 residues. The partial region of HK97 gp5 is shown from residue 176 to 180, and from 226 to 234 (T loop). The HK97 sequences are written underlined in color blue. The protein residues corresponding to the HK97 residue K178 and D231 are written in color red.

### **The model of HK97 prohead formation, assembled from the capsomers**

My data reveal that residue 178 pairs with residue 231 to assemble the HK97 proheads, and the mutational analysis results imply that D231A and D231L cannot interact with residue K178 as the Ala-Lys (A-K) and Leu-Lys (L-K) interactions for assembling capsomers into proheads. Nevertheless, further results from the mutational analysis of residue 178 showed that two single mutations K178L and K178E cannot interact with residue D231 as the Asp-Leu (D-L) and Asp-Glu (D-E) interactions to assemble capsomers into proheads.

My data strongly demonstrate that the HK97 proheads cannot be assembled without the correct pair of residues 231 and 178 between capsomers. The second-site suppressors K178V, K178I, and K178N in the isolated pseudo-revertants were found to fully suppress the lethal mutation D231L in gp5. Either a salt bridge or a hydrophobic interaction can allow assembling capsomers into proheads.

Since the difference between the wild type and three revertants is the two single substitutions in gp5, the above data imply that the functional pairs of the residues 231 and 178 are Asp-Lys (D-K), Leu-Val (L-V), Leu-Ile (L-I), and Leu-Asn (L-N), but not the pairs of the Ala-Lys (A-K), Leu-Lys (L-K), Asp-Leu (D-L), and Asp-Glu (D-E). This clarifies that HK97 can produce viable bacteriophages when the correct two amino acids are presented at residues 231 and 178 in gp5.

To summarize my study of residues 231 and 178 in HK97, the interaction of residues 231 and 178 plays a central and crucial role in assembling the HK97 capsomers into proheads. Interestingly, the difference between the wild type and the revertants is only two single substitutions in the gp5. The pair of residues 231 and 178 is not required for subunit gp5's folding and the pathway of assembling the gp5's into capsomers, but is required for assembling

capsomers into proheads. However, the discovery of one function of a residue does not rule out the possibility of other functions of that residue. My study provides convincing evidence that residues 231 and 178 are essential for assembling prohead, but does not determine whether these residues are involved in other maturation pathways. But there is no doubt that this special interaction between residues 231 and 178 offers an essential step for the bacteriophage HK97 capsid assembly. I propose that one of the major functions of residues 231 and 178 is to provide the correct T=7 capsid assembly of the bacteriophage HK97. Either the ionic (*e.g.* salt bridge) or the hydrophobic interactions of residues 231 and 178 can support HK97 capsid assembly.

## 6.0 CONCLUSIONS

Amino acids have functional roles in protein stability, activity, and specificity (Parent, Ranaghan, and Teschke, 2004; Schoepp and Johnston, 1993; Zhang et al., 2009). The bacteriophage HK97 capsid is assembled from hundreds of copies of major capsid protein gp5, so the amino acid residues of gp5 must have their particular roles in the serial stages of capsid assembly and maturation. Studying the single mutations that interfere with capsid formation is a way of identifying crucial gp5 residues. However, these mutants cannot provide enough information about the exact residue functionality. Therefore, isolating the revertants that contain the suppressors is a better method for identifying the functional roles of residues, which also provide the information about the networks of interaction within and between subunits on the HK97 capsid. Here I use genetic, biochemical, and structural approaches to characterize the roles of the gp5 residue 163 in HK97 head protein crosslinking formation. I also determined the functions of residues 231 and 178 in HK97 capsid assembly.

## **6.1 ISOLATING HK97 REVERTANTS FOR IDENTIFYING SUPPRESSORS IS FEASIBLE**

In this report, the genetic technique for isolating suppressors is developed to investigate the protein interactions on the bacteriophage HK97 capsid. By using this approach, any lethal mutation can be introduced in the essential gene 5 to generate a detrimental phenotype for isolating revertants. These isolated revertants bearing the suppressors have provided useful genetic data *in vivo*. The tolerated amino acids were isolated at residue 163 (Chapter 4), and the second-site suppressors were isolated at residue 178 to suppress the lethal mutation at residue 231 (Chapter 5).

In chapter 4, I demonstrated that mutant V163D exhibits a defect in the head protein crosslinking; the major one may be a failure to provide a hydrophobic environment for triggering the HK97 self-crosslinking. In chapter 5, I reported that single substitutions at residues 231 (or 178) impair the assembly of capsomers into proheads and the defect can be corrected by the alternative pairs between residues 231 and 178. In both cases, identification of suppressors underlined the functional importance of the HK97 gp5 residues and clarify the nature of their functional roles.

The mutant lysogen DT-V163D was the first mutant lysogen that was induced for obtaining pseudo-revertants. It was found that DT-V163D most readily produced true revertants, V163V, since the true revertant V163V (GTT) only requires one substitution. However, there was one pseudo-revertant isolated in the four revertants from my study, indicating that the

isolation of the pseudo-revertants from the mutant lysogens is not very difficult. It was very exciting to identify the second-site suppressors K178V, K178I, and K178N in the revertants after isolating the revertants from mutant lysogens DT-D231L, assuring that isolating suppressors in HK97 structural proteins is practical.



## 6.2 THE LIMITATION OF THE GENETIC APPROACH

Suppressor isolation is a powerful genetic approach for identifying side-chain interactions. Since the selection is based on phage HK97 viability, only the revertants which contain functional gp5's can be isolated in my study. However, some of my constructed mutant lysogens give no surviving phage, such as DT-E165R, DT-E165L, DT-R365D, and DT-N278R (Table 1 and data not shown). The possible reason may be their low reversion frequency (one in  $>10^{11}$ ) (see below).

The lethal mutations in HK97 gp5 may change back to the original codons, and the true revertants may be the most readily isolated revertants. A true revertant has been isolated from the mutant glycoprotein E1 Y472S in the Rubella virus (Yao and Gillam, 2000). The substituted Ser was reverted to the original residue Tyr (Yao and Gillam, 2000). Thus, there is not much information obtained from these true revertants. In my study, true revertants were the only survivors isolated from the lysogens DT-L224Q and DT-E165K.

The residue E165 is conserved in the major capsid proteins among the dsDNA bacteriophages' major capsid proteins (Figure 18, Appendix B). The single substitutions E165R, E165K, E165L, and E165A were tested and determined to be lethal mutations (Appendix A). Residue E165 is located at the outer end of the gp5 E loop and the E loop translocates during prohead expansion. Thus, residue 165 may play an important role in HK97 capsid assembly and maturation. In order to investigate the functions of residue 165, mutant lysogens DT-E165R, DT-E165L, and DT-E165K were constructed and induced for isolating the pseudo-revertants. No revertant was isolated from the above mutant lysogens except for DT-E165K. Four true

revertants were isolated from DT-E165K and the reversion frequency is one in  $8.56 \times 10^{10}$ . The suppressors were determined to be E165E (GAG), changed from the single substitution E165K (AAG), suggesting that the mutant lysogen DT-E165K most readily produces the true revertant E165E. These data indicate that two nucleotide substitutions may be required for avoiding isolating true revertants.

The data from the mutant lysogens E165R and E165L showed that some of the mutant lysogens have a difficulty to produce any revertant, suggesting that these suppressors may occur at a very low frequency and cannot be isolated in my study. A large scale lysogen culture or a different mutation designed in the lysogen may be necessary for future experiments. The reversion frequencies may vary among different mutant lysogens. For instance, the reversion frequency of obtaining the true revertants E165E is one in  $\sim 10^{11}$  and the reversion frequency of the true revertants L224L is around one in  $\sim 10^8$  (see below). Thus, it is difficult to predict these *in vivo* results. Furthermore, it is interesting to know that the reversion frequency of the HK97 gp5 in my study is much lower than that from the Salmonella phage P22 tailspike protein (typically between  $10^{-4}$  and  $10^{-7}$ ) (Villafane, Fleming, and Haase-Pettingell, 1994) and from the *E.coli* bacteriophage  $\Phi$ X174 (range from  $10^{-3}$  to  $10^{-7}$ ) (Yao and Gillam, 2000). The reason is unclear.

The gp5 residue L224 is conserved among dsDNA bacteriophages (Appendix B) and the single substitutions L224P, L224R, and L224Q are lethal mutations (Appendix A). To determine the functional roles of residue 224, the mutant lysogen DT-L224Q was induced for identifying suppressors. Three true revertants were isolated and the suppressors in the revertants were determined to be L224L (CTA), mutated from L224Q (CAA). This indicates that mutant lysogen DT-L224Q most readily produces the true revertant L224L. All true revertants carry the codon

CTA rather than the original codon CTG since the change from Gln (CAA) to Leu (CTA) only requires one nucleotide substitution. This finding is in agreement with the observation that the codon used for generating a mutation requires two nucleotide substitutions to avoid getting true revertants.

On the other hand, some lysogens constructed with two (or more than two) single nucleotide substitutions cannot produce any revertant at all. For instance, the mutant lysogens DT-E165R, N278R, and R365D were constructed with more than two nucleotide substitutions in the gp5, but there was no revertant isolated from any of them. The reason is unknown, but one possible reason may be that two nucleotide substitutions are difficult for these lysogens to generate the reversion. Besides, some of the gp5 residues may have complicated roles in the HK97 capsid assembly and maturation and may require a reversion with more than one kind of suppressors, making it difficult to gain revertants.

### 6.3 GENETICS-BASED PROTEIN ANALYSIS

The second-site suppressors have been isolated in some of the phage (or virus) structural proteins in the literatures (Murialdo and Tzamtzis, 1997; Parent, Ranaghan, and Teschke, 2004; Villafane, Fleming, and Haase-Pettingell, 1994; Vlaycheva et al., 2004). However, the functions of these suppressors and the meanings of obtaining these suppressors are rarely known. In my research, genetic suppressor study was correlated with the biochemical and structural analysis to identify the functionality of the HK97 gp5 residues, e.g. V163 and D231. The crystal structures of the HK97 Head II and Prohead II have been determined and are available in the protein databases, providing an overall protein network on the HK97 capsid. So the global influence of a single suppressor substitution on HK97 capsid can be determined. For instance, the proper residue 178 was not only able to support the local protein interaction with residue 231, but also rescue the HK97 prohead assembly from capsomers.

Rapid accumulation of the structural data provides lots of protein information in the protein databases, which are determined in the past couple years expeditiously. What would come next after determining the protein structures? How could we use this structural information? The genetic data from suppressors may be helpful. As in my research, the biological importance of the gp5 residue interactions and the protein networks on the HK97 capsid can be determined by a concert of *in vivo* genetic data and the visualization of protein structures.

## APPENDIX A

### THE BIOLOGICAL ACTIVITIES OF HK97 GP5 MUTANTS

(Data were collected by Dr. Li, Dr. Baros, and Danju Tso)

#### **Table 9: The summary of biological activities of HK97 gp5 mutants**

The single residues in the HK97 gp5 have been substituted by either an Ala or other amino acids. The biological activities of the mutant gp5's were determined by complementation spot assay (see Chapter 2, Method). The mutant gene 5 was constructed in the plasmid pV0 for examining the ability of complementing the gene 5<sup>-</sup> amber phage. The biological activity of wild type gp5 was determined and counted as 100%. The biological activity of mutant gp5 was counted for <0.01% when mutant shows <10000 fold less than the wild type. The following table compiles the single mutations in HK97 gene 5 that have been characterized by complementation spot assay in this study and in earlier studies in our laboratory. Data were summarized by Danju Tso.

Residue	Mutation	Result (%)
<b>Wild type</b>	–	<b>100</b>
S107	S107K	0.1
D108	D108E	10
	D108A	1
N146	N146G	100
	N146R	100
	N146L	100
E149	E149K	<0.01
V163	V163D	<0.01
E165	E165D	<0.01
	E165K	<0.01
	E165A	<0.01
	E165R	<0.01
	E165L	<0.01
K169	K169Y	<0.01
	K169N	<0.01
K178	K178L	<0.01
	K178E	<0.01
Q191	Q191L	1
	Q191R	1
D199	D199L	<0.01
	D199N	<0.01
M202	M202I	<0.01
L203	L203P	1
Q204	Q204P	<0.01
S205	S205P	<0.01
I207	I207T	100
N208	N208Y	1
	N208S	100
N209	N209Y	100
L211	L211P	<0.01
M212	M212L	100
E219	E219K	<0.01
	E219G	100
E220	E220G	<0.01
L223	L223P	<0.01
	L223Q	<0.01
L224	L224R	<0.01
	L224P	<0.01

	L224Q	<0.01
D231	D231L	<0.01
	D231A	<0.01
Deletion of loop	Δloop	<0.01
G235V	G236	<0.01
L236	L236P	<0.01
	L236R	<0.01
N237	N237D	1
A240	A240T	<0.01
Y243	Y243C	100
	Y243N	100
D244	D244G	1
L247	L247P	<0.01
	L247Q	100
N248	N248Y	1
I258	I258T	1
I262	I262N	<0.01
Q264	Q264P	<0.01
	Q264R	<0.01
N278	N278L	<0.01
	N278R	0.1
R280	R280I	1
D351	D351N	10
	D351Q	0.1
N356	N356D	<0.01
E363	E363D	<0.01
	E363N	<0.01
	E363H	<0.01
	E363A	<0.01
	E363Y	<0.01
	E363Q	<0.01
R365	R365D	<0.01
	R365K	<0.01
	R365N	<0.01

## APPENDIX B

### THE SEQUENCE ALIGNMENT OF THE DSDNA BACTERIOPHAGES' MAJOR CAPSID PROTEINS

The major capsid proteins of nine dsDNA bacteriophages, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B, have similarity to *E. coli* bacteriophage HK97 gp5 (385a.a.). The nine bacteriophages are *Pseudomonas* D3 (395a.a.), *Xanthomonas* phage OP1 (390a.a.), *Xanthomonas* phage Xp10 (390a.a.), *Xanthomonas* phage Xop411 (390a.a.), *Streptomyces* phage Mu1/6 (419a.a.), *Flavobacterium* phage 11b (379a.a.), *Corynebacterium* phage BFK20 (413a.a.), Enterobacteriophage SFV (409a.a.) and *Burkholderia* phage phi1026b (418a.a.). The sequences were analyzed by ClustalW. The HK97 sequences are written in color blue and the HK97 T loop region is underlined. The residues corresponding to the HK97 residue V163 are written in color red. The residues corresponding to the HK97 residue K178 are written in color pink. The protein residues corresponding to the HK97 residue D231 are written in color green.



HK97	MSELAL-----IQKAI EESQQKMTQLFDAQKAEIESTGQVSKQLQ	40
D3	MSDFEK--QIG-----ELNASLKQVGDQIKSQAEQVNTQIANFGEMNKETR	44
OP1	MTDITA-----KLEATLANVTDSLKA FGERAVRD----GELNASAR	37
XP10	MTDITS-----KLEATLANVTDSLRAFGERAVRD----GELNASAR	37
XOP411	MTDITS-----KLEATLANVTDSLRAFGERAVRD----GELNASAR	37
MU1/6	MPPTPT-----LEEQRAALLARLDDTSLTTEQVQEI VAEARGLADALQ	43
11B	MEALEIK-----VALEAIKGQVDSKSSAQALEVKGLIEALEAKMTSE	42
BFK20	MVKEAG-----DAPTNAQVAEIAEVKSMVEQFKADEDAKRERAKSVK	42
SFV	MKLHELK-----QKRNTIATDMRALNEKIGDNAWTEEQRTTEWNK	39
PHI1026B	MSHMNEPRQFGRKSGGDSHPQVLETVTVKELKRIGDEVKSAGEKALAEAKRAGDLGVETK	60
HK97	SDLMKVQEELTKSGTRLFDLEQK--LASGAENPGEKKSFSERAAEELIKS--WDGKQGTG	96
D3	AKVDELLTAQGELQARLSAAEQ--MLANEK-RDGGEEAPKTAGQMVAESLKEQGVTS	101
OP1	SKYDELFATVGNLSAEVQAARQR--VAELEGNGAGGDVQHVSVDGMFVASEQFQASTGRW	95
XP10	SKYDELFATVGNLSAEVQAARQR--VAELEGNGAGGDVQHVSVDGLFVASEQFQASAGRW	95
XOP411	SKYDELFATVGNLSAEVQAARQR--VAELEGNGAGGDVQHVSVDGMFVASEQFQASAGRW	95
MU1/6	AESDRAAARAALLRTAPPAPKGP--ADGGTPLTPAEAGTFRSLAQRFA DSDGLREYRARD	101
11B	KDLAVNELKSDMAALQAHADKLDVKLKEKAKSEDKSDSLVKSITENFN DIKEVRNGKSIQ	102
BFK20	ANQDFLRELQEATAGSV DSEKSGELTRKGEGYKSI GEFFAKRAGDQIKQAGGAQLNYSV	102
SFV	AKSELEALDERIAREEELRRQDQAYIESNEEEQRQNLDPENNSQQDEKRAQVFDKWMRHG	99
PHI1026B	ATVDELLIKQGELQARLLEAEQ--KLARGG-GSAELET PKTLGQLVTESEEMKGMDS	116
HK97	G-----AKTFNKSLGSDADSAG-----SLIQPMQIPGIIMP-GLRRLTIRDLLAQ	141
D3	R-GSHRVSMPSAITSIDGSGG-----ALVAPDRRPGVVAA-PQRRLTIRDLVAPG	150
OP1	NDRSARATMNIKAALNTASTDAAGSAG----ALTPNRLPGFITP-PDARLTVRDLIGSG	150
XP10	NDRSARATMNIKAALNTASTDAAGSAG----ALTPNRLPGFITQ-PDARLTVRDLIGSG	150
XOP411	NDRSARATMNIKAALNTASTDAAGSAG----ALTPNRLPGFITP-PDARLTVRDLIGSG	150
MU1/6	KRGQFQVEMRDI DPNRLLSRDAPAGTITPNVPHLPQLVPGIVPTT PDLPLLVADLLDQQ	161
11B	VKAVGDMTLPVNL TGAQP-----KDYNFDVVLN-PSQMLNVSDIVGAV	144
BFK20	GEYVAPRVKAASDPASTATLTDEFQGG-----YGTTWNRNI IYR-RREKL VVADLMDNL	155
SFV	ASELTSEERKALRELRAQGV AQDEKGG----YTVPETFLAKVVERMKSYGGIASVAQILT	155
PHI1026B	R-KSVRVVRDRKSI MNVPATVSGSGVSGS--NSLVVADRQAGI IAP-PQRKMTIRDLLMPG	172

HK97 RTSSNALEYVREEV-----FTNNADVVAEKALKPESDIT-FSEKQTA-NVKTIAHVVQA 192  
D3 TTESNSVEYVRETG-----FVNNAAPVSEGTQKPYSDLT-FELENA-PVRTIAHLFKA 201  
OP1 RTDSALIEYVQETG-----FVNNAAIVAEGALKPESSLK-FAKKTD-TTHVIAHTMKA 201  
XP10 RTDSALIEYVQETG-----FVNNAAIVAEGALKPESSLK-FAKKTD-TTHVIAHTMKA 201  
XOP411 RTDSALIEYVQETG-----FVNNAAIVAEGALKPESSLK-FAKKTD-TTHVIAHTMKA 201  
MU1/6 NADYNVLEYIRDTSGTAGAGSTWNKAAVVEGTAKPQSTLS-FDTITT-TLKTVAHWLPI 219  
11B SISGGTYTFVRENGAG-----EGAIGAQVEGATKGQKDYD-ISMIDV-NTDFIAGFTRY 196  
BFK20 TMTNTTIKYLMEKAN---RVVEGGFKTVAEGGKKPYMRFAFDI VTE-SLSKIAGLTKI 210  
SFV TSDGRTEWATADG-----TSEVGVLLGENEEAGEEDTDFGNGSLGALKMTSKI IRV 207  
PHI1026B QTSSSSIEYTVETG-----FTNNAAA VEGAQKPTS DLK-FNLKNQ-PVRTIAHLFKA 223

HK97 SRQVMDAPM-LQSYINNRLMYGLALKEEGQLLNGDG--TGDNLEGLNKVATAYDTS--L 247  
D3 SRQILDDASA-LQSYIDARARYGLMLVEECQLLYGNG--TGANLHGIIPQAQAYAPPSGV 258  
OP1 TRQILSDAPQ-LASYMNNRLIRGLVKEDAEILRGTG--ANDGLLGLIPQATTYAAP--T 256  
XP10 TRQILSDAPQ-LASYMNNRLIRGLVKEDAEILRGTG--ANDGLLGLIPQATTYAAP--T 256  
XOP411 TRQILSDAPQ-LASYMNNRLIRGLVKEDAEILRGTG--ANDGLLGLIPQATTYAAP--T 256  
MU1/6 TRQAADDNSQ-LMGYIQGRITYGLRFLRDRQLLNGNGSTEMQGILTTPIGTYYQPKPTA 278  
11B SKKMANNLPF-LTSFIPNALRRDYAKAEN-----AAFNVLAAANATASTE 240  
BFK20 TDEMIEDYDF-LVSYINARLLEELAEIEERQLLLGDG--TGNNLTGLLKRDIQTALA--V 265  
SFV SNELLQDSAIDMEAYLARRIAERIGRGEARYLIQGTGAGTPKQPKGLAASVTGTTQTA 267  
PHI1026B SRQILDDAPA-LQSYIDGRARYGLQLTEEGQILKGDG--TGANILGILPQASAFMP--SI 278

HK97 NATGDTRADIIAHAIYQVTESEFSASGIVLNPRDWHNIALLKDNAG--RYIFGGPQA--- 302  
D3 VVTAEQRIDRIRLAILQAQLAEFPASGIVLNPIDWALIELNKDAEN--RYIIGSPQN--- 313  
OP1 TIAGATRVQDLRLAMLQASLAEPASGIVINPIDWAAIELAKDANN--QYLIGNARG--- 311  
XP10 TIAGATRVQDLRLAMLQASLAEPASGIVINPIDWAAIELAKDANN--QYLIGNARG--- 311  
XOP411 TIAGATRVQDLRLAMLQASLAEPASGIVINPIDWAAIELAKDANN--QYLIGNARG--- 311  
MU1/6 PATDEPPLVDIRRAKTVAEIAFGPPDGVVVHPQDWESI ELQAPGSGVFRVIANVQG--- 335  
11B IITNKNKVEMLINEIAKQENLDFPVTAIVLRPTDYYDILVTQKSVGAGYGLPGVVTQD-- 298  
BFK20 SNKDELADSIYKAMTNISLATPFQADALVINPLDYQELRLAKDANG--QYYGGGVFQGGY 323  
SFV NAVKWQEILALKHSIDPAYRRG-PKFRALFNDNTLKLISEMEDGQGR-PLWLPIVIG--- 322  
PHI1026B TLANATPIDKIRLALLQAVLAEPATGIVLNPIDWASIELTKDSQG--RYIVGNPNVN--- 333

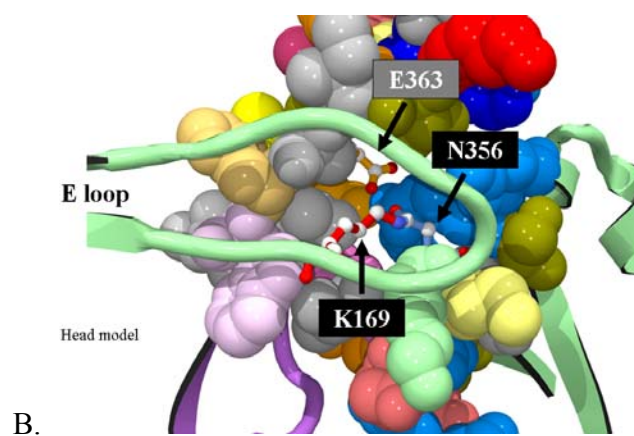
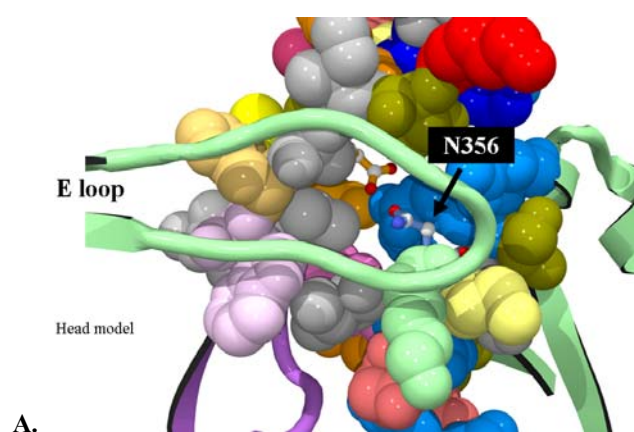
HK97	-----FTSNIMWGLPVVPTKAQAAG-----TFTVGGFDMASQVWDRMDATVEVSREDRDN	352
D3	-----GTTPTLWRLPVVETQAITQD-----EFLTGAFLSLGAQIFDRMDIEVLVSTENDKD	363
OP1	-----TLTPTLWGLPVVATQAMAPG-----EFLVGAFDLAAQIFDQWDARVEIGYVN-DD	360
XP10	-----TLTPTLWGLPVVATQAMAPG-----EFLVGAFDLAAQIFDQWDARVEIGYVN-DD	360
XOP411	-----TLTPTLWGLPVVATQAMAPG-----EFLVGAFDLAAQIFDQWDARVEIGYVG-ED	360
MU1/6	-----EATPRIWGLNVVSTVAIAQG-----TALVGGFRQGATLWSRQGITVLMTDSHADP	385
11B	-----NGVLRINGIPLFRATWLAAN-----KYYVGDWTRVTKVTTEG-LSLEFSEVEGTN	347
BFK20	GSGGIMLDPAPWGLRTVQSQVVPVG----KPVVGAFRSAASVLRKGGVRIDSTNTNVDD	378
SFV	-----VAPASVLNVPPYVIDQEIDDIGAGKKFMFCGDFDRFIIRRVRYMILKRLVERYAEY	377
PHI1026B	-----GTTPLWNLPPVETQAMTAN-----EFLVGAFSMAAQIFDRMEIEVLLSTENVDD	383
HK97	FVKNNMLTILCEERLALAHYRPTAIKGTFFSSGS--	385
D3	FENNMVTIRAEERLAFVYRPEAFVTGSLTAS---	395
OP1	FQRNMVTVLAEERLALVYRPEALITGSFA-----	390
XP10	FQRNMVTVLAEERLALVYRPEALISGSFA-----	390
XOP411	FQRNMVTVLAEERLALVYRPEALISGSFA-----	390
MU1/6	FTANTLVILAEFRANLAVYQPKAFVRVTFAAATT-	419
11B	FVKNNITARIEAQVALAVEQPAALIFGDFTAV---	379
BFK20	FENNLITVRAEERVGLMVTTFPEAIVQLDVAEVVTP	413
SFV	DQTGFLAFHRFDCILEDTSAIKALVGKGSVGG---	409
PHI1026B	FEKNMVSIRAEERLALAVYRPEFVTGALVEQAGG	418

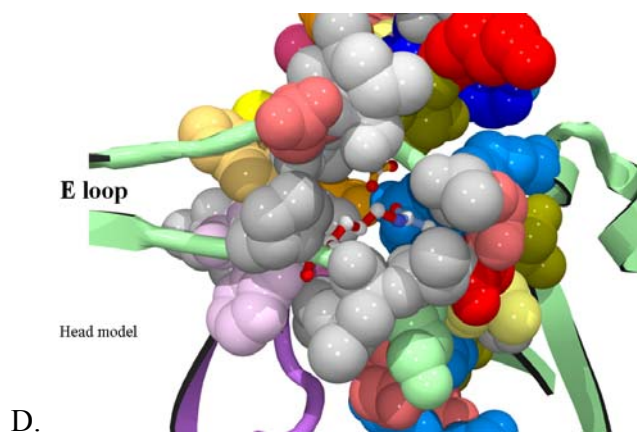
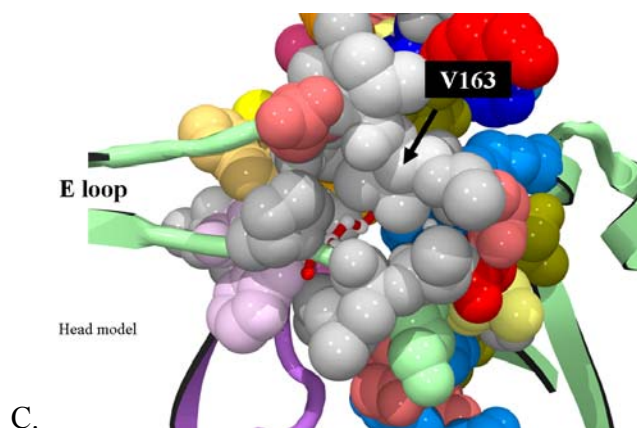
**Figure 43: The major capsid protein sequence alignment of the dsDNA bacteriophage HK97, D3, OP1, XP10, XOP411, MU1/6, 11B, BFK20, SFV and PHI1026B**

## APPENDIX C

### THE HK97 HEAD CROSSLINK SITE

(Pictures are provided by Dr. Robert Duda)





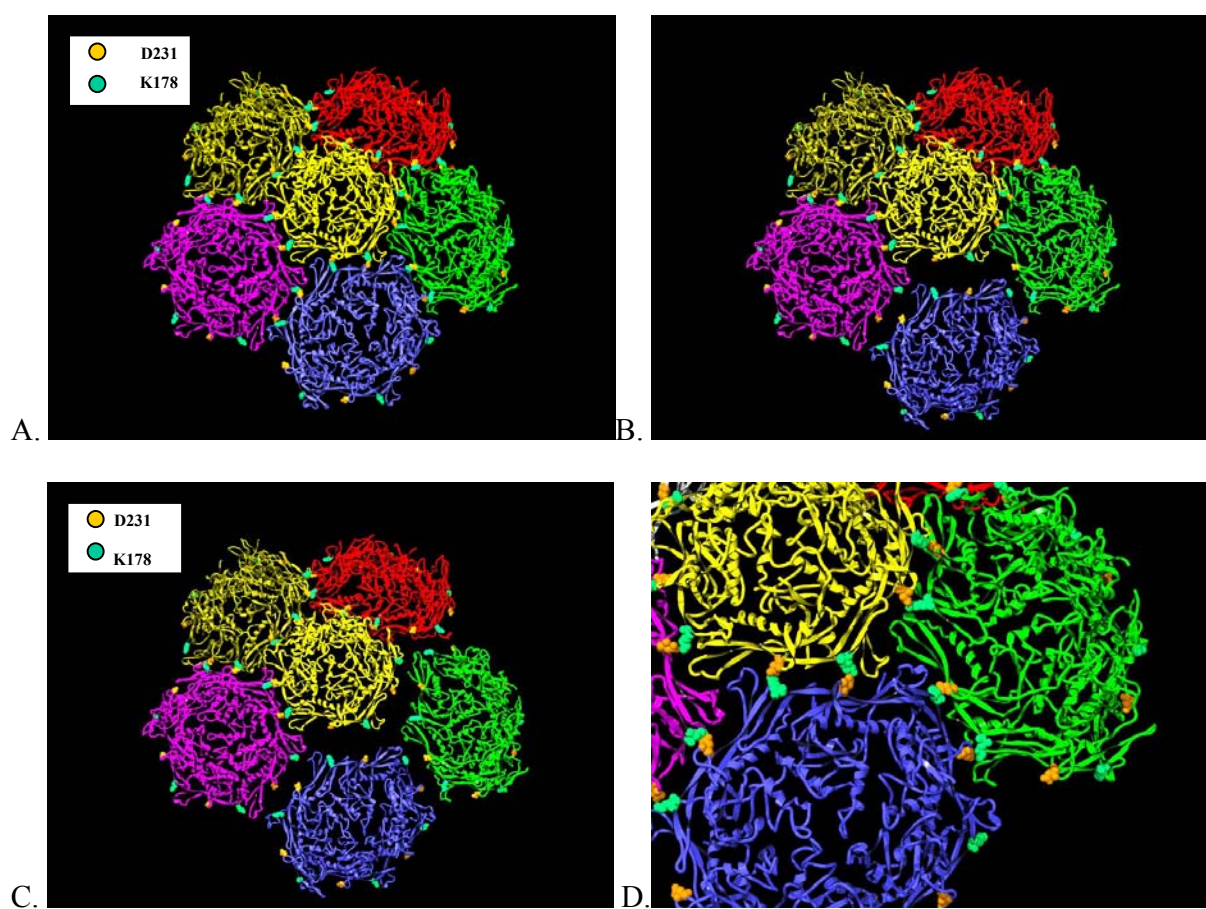
**Figure 44: The HK97 head crosslink site**

The side chains of residue N356 is labeled in Panel A. The side chains of the residues N356, K169, and E363 are labeled in Panel B. Panel A and B are without the local residues in the E loop. Panel C and D are with the local residues in the E loop except for the residue V163 in Panel D. The residue V163 is labeled in Panel C.

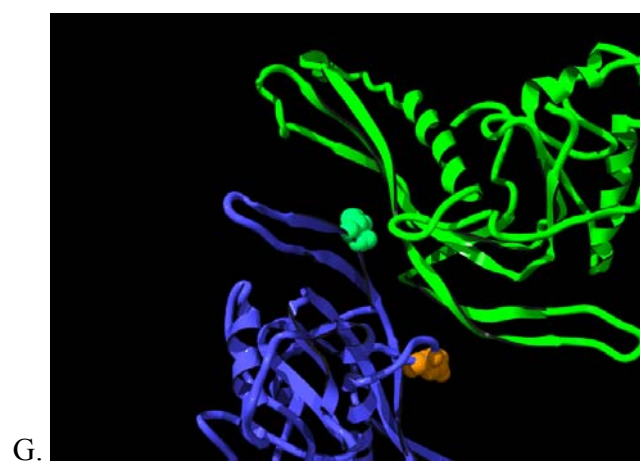
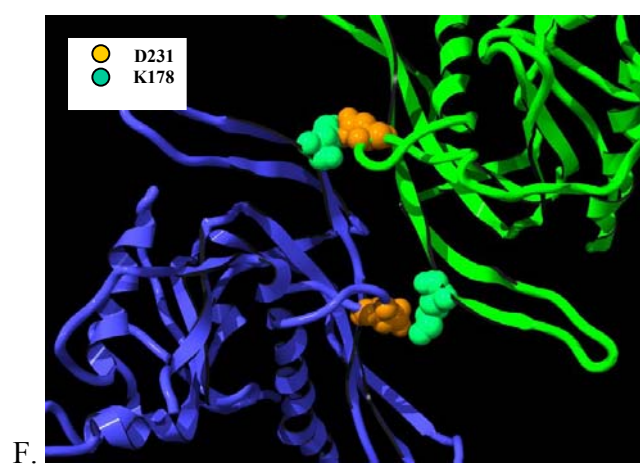
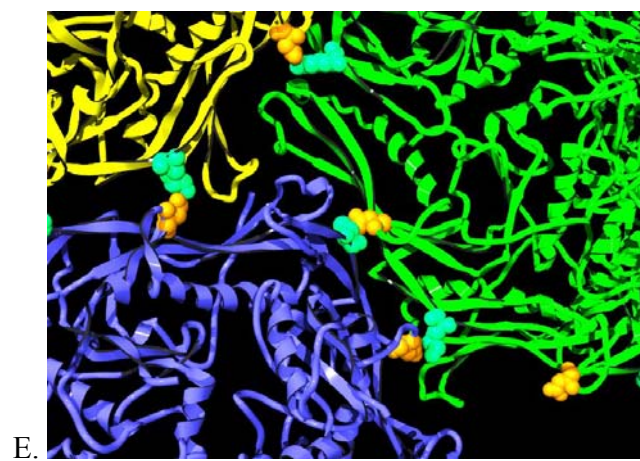
## APPENDIX D

### THE LOCATION OF THE RESIDUES D231 AND K178 ON HK97 PROHEAD II

(Pictures are provided by Dr. Robert Duda)







**Figure 45: The HK97 prohead assembly**

The partial Prohead II is shown in Panel A, B and C, including one pentamer and five hexamers. The pentamer is located at the center region surrounded by five hexamers. The residues D231 and K178 face out of each capsomer instead of being buried inside the capsomers. Panel D, E, F, and G are the close-up pictures of the pentamer-hexamer-hexamer junction. The relationship of the residues D231 and K178 can be found either between pentamers and hexamers, or between hexamers and hexamers. Panel F shows one junction between two nearby hexamers. Each residue D231 interacts to the residue K178, which is located at the neighboring gp5. Panel G is the same view of panel F but without one pair of the residues D231 and K178.



## APPENDIX E

### THE PRIMERS FOR AMPLIFYING AND SEQUENCING THE HK97 GENE 5

**Table 10: The list of primers**

Primer name	Direction	Sequence
DRHPL3	Forward	CGTAAAAACGCTGTCCTTCG
LP1	Forward	GGTGACGGGAAGCAGGGG
GP4-F1*	Forward	CCAACGAACAGGCTGGCATCGC
GP4-R1*	Reverse	CGCTGACTTAAACCGGGCTA
MP4	Forward	GAGCGAGTCCGAGGGCG
96-45	Forward	AGCAGAAATCGAAAGCACAG
RMP3	Forward	CCTTCAGCAAACAAACCG
SP10	Forward	CGCCTATGACACCTCGC
RMP4	Reverse	GGG TTCAGGACGATACCG
SP9	Reverse	GCGCCAGACGCTCTTC
RMP5	Reverse	CCCCTCCATCATGAGCC
GP6-R1*	Reverse	GTCGTCCCTGTCGTCTTCCTC
GP6-R2*	Reverse	GAATGAACATCATGCGTTCGGCG
GP7-R1*	Reverse	CCCTGTTCCTGAATCGTCCCGGTTCC

The primers labeled with\* were designed by Danju Tso

## BIBLIOGRAPHY

- Ackermann, H. W. (1992). Frequency of morphological phage descriptions. *Arch Virol* **124**(3-4), 201-9.
- Ackermann, H. W., and Kropinski, A. M. (2007). Curated list of prokaryote viruses with fully sequenced genomes. *Res Microbiol* **158**(7), 555-66.
- Adereth, Y., Champion, K. J., Hsu, T., and Dammai, V. (2005). Site-directed mutagenesis using Pfu DNA polymerase and T4 DNA ligase. *Biotechniques* **38**(6), 864, 866, 868.
- Aramli, L. A., and Teschke, C. M. (1999). Single amino acid substitutions globally suppress the folding defects of temperature-sensitive folding mutants of phage P22 coat protein. *J Biol Chem* **274**(32), 22217-24.
- Bazinet, C., Villafane, R., and King, J. (1990). Novel second-site suppression of a cold-sensitive defect in phage P22 procapsid assembly. *J Mol Biol* **216**(3), 701-16.
- Beissinger, M., Lee, S. C., Steinbacher, S., Reinemer, P., Huber, R., Yu, M. H., and Seckler, R. (1995). Mutations that stabilize folding intermediates of phage P22 tailspike protein: folding in vivo and in vitro, stability, and structural context. *J Mol Biol* **249**(1), 185-94.
- Bellingham, C. M., Lillie, M. A., Gosline, J. M., Wright, G. M., Starcher, B. C., Bailey, A. J., Woodhouse, K. A., and Keeley, F. W. (2003). Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. *Biopolymers* **70**(4), 445-55.
- Berget, P. B., and Poteete, A. R. (1980). Structure and functions of the bacteriophage P22 tail protein. *J Virol* **34**(1), 234-43.
- Borriess, M., Lombardot, T., Glockner, F. O., Becher, D., Albrecht, D., and Schweder, T. (2007). Genome and proteome characterization of the psychrophilic *Flavobacterium* bacteriophage 11b. *Extremophiles* **11**(1), 95-104.
- Bowzard, J. B., Wills, J. W., and Craven, R. C. (2001). Second-site suppressors of Rous sarcoma virus Ca mutations: evidence for interdomain interactions. *J Virol* **75**(15), 6850-6.
- Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F., and Schneewind, O. (2008). Amide bonds assemble pili on the surface of bacilli. *Proc Natl Acad Sci U S A* **105**(29), 10215-20.
- Bukovska, G., Klucar, L., Vlcek, C., Adamovic, J., Turna, J., and Timko, J. (2006). Complete nucleotide sequence and genome analysis of bacteriophage BFK20--a lytic phage of the industrial producer *Brevibacterium flavum*. *Virology* **348**(1), 57-71.
- Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F., and Hendrix, R. W. (2005). The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol* **187**(3), 1091-104.
- Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. L., and Brussow, H. (2004). Phage-host interaction: an ecological perspective. *J Bacteriol* **186**(12), 3677-86.
- Cimarelli, A., Sandin, S., Hoglund, S., and Luban, J. (2000). Rescue of multiple viral functions by a second-site suppressor of a human immunodeficiency virus type 1 nucleocapsid mutation. *J Virol* **74**(9), 4273-83.

- Claverie, J. M., de Souza, A. C., and Thirion, J. P. (1979). Mutations of Chinese hamster somatic cells from 2-deoxygalactose sensitivity to resistance. *Genetics* **92**(2), 563-72.
- Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W., and Steven, A. C. (1995). Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. *J Mol Biol* **253**(1), 86-99.
- Conway, J. F., Wikoff, W. R., Cheng, N., Duda, R. L., Hendrix, R. W., Johnson, J. E., and Steven, A. C. (2001). Virus maturation involving large subunit rotations and local refolding. *Science* **292**(5517), 744-8.
- Copeland, N. G., Jenkins, N. A., and Court, D. L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**(10), 769-79.
- Court, D. L., Sawitzke, J. A., and Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88.
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**(12), 6640-5.
- Datta, S., Costantino, N., and Court, D. L. (2006). A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109-15.
- Debouck, C., Riccio, A., Schumperli, D., McKenney, K., Jeffers, J., Hughes, C., Rosenberg, M., Heusterspreute, M., Brunel, F., and Davison, J. (1985). Structure of the galactokinase gene of *Escherichia coli*, the last gene of the gal operon. *Nucleic Acids Res* **13**(6), 1841-53.
- Dhillon, E. K., Dhillon, T. S., Lai, A. N., and Linn, S. (1980). Host range, immunity and antigenic properties of lambdoid coliphage HK97. *J Gen Virol* **50**(1), 217-20.
- Dierkes, L. E., Peebles, C. L., Firek, B. A., Hendrix, R. W., and Duda, R. L. (2009). Mutational analysis of a conserved glutamic acid required for self-catalyzed cross-linking of bacteriophage HK97 capsids. *J Virol* **83**(5), 2088-98.
- Duda, R. L. (1998). Protein chainmail: catenated protein in viral capsids. *Cell* **94**(1), 55-60.
- Duda, R. L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D., and Hendrix, R. W. (1995a). Structural transitions during bacteriophage HK97 head assembly. *J Mol Biol* **247**(4), 618-35.
- Duda, R. L., Martincic, K., and Hendrix, R. W. (1995). Genetic basis of bacteriophage HK97 prohead assembly. *J Mol Biol* **247**(4), 636-47.
- Duda, R. L., Martincic, K., Xie, Z., and Hendrix, R. W. (1995b). Bacteriophage HK97 head assembly. *FEMS Microbiol Rev* **17**(1-2), 41-6.
- Effantin, G., Boulanger, P., Neumann, E., Letellier, L., and Conway, J. F. (2006). Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J Mol Biol* **361**(5), 993-1002.
- Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**(12), 6742-6.
- Falker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008). Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol* **70**(3), 595-607.
- Fane, B., and King, J. (1991). Intragenic suppressors of folding defects in the P22 tailspike protein. *Genetics* **127**(2), 263-77.

- Fane, B., Villafane, R., Mitraki, A., and King, J. (1991). Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *J Biol Chem* **266**(18), 11640-8.
- Fane, B. A., and Hayashi, M. (1991). Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage phi X174. *Genetics* **128**(4), 663-71.
- Fokine, A., Chipman, P. R., Leiman, P. G., Mesyanzhinov, V. V., Rao, V. B., and Rossmann, M. G. (2004). Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A* **101**(16), 6003-8.
- Fokine, A., Leiman, P. G., Shneider, M. M., Ahvazi, B., Boeshans, K. M., Steven, A. C., Black, L. W., Mesyanzhinov, V. V., and Rossmann, M. G. (2005). Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Natl Acad Sci U S A* **102**(20), 7163-8.
- Fong, D. G., Doyle, S. M., and Teschke, C. M. (1997). The folded conformation of phage P22 coat protein is affected by amino acid substitutions that lead to a cold-sensitive phenotype. *Biochemistry* **36**(13), 3971-80.
- Gan, L., Conway, J. F., Firek, B. A., Cheng, N., Hendrix, R. W., Steven, A. C., Johnson, J. E., and Duda, R. L. (2004). Control of crosslinking by quaternary structure changes during bacteriophage HK97 maturation. *Mol Cell* **14**(5), 559-69.
- Gan, L., Speir, J. A., Conway, J. F., Lander, G., Cheng, N., Firek, B. A., Hendrix, R. W., Duda, R. L., Liljas, L., and Johnson, J. E. (2006). Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure* **14**(11), 1655-65.
- Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., Hendrix, R. W., Komives, E. A., and Johnson, J. E. (2009). An unexpected twist in viral capsid maturation. *Nature* **458**(7238), 646-50.
- Gilakjan, Z. A., and Kropinski, A. M. (1999). Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? *J Bacteriol* **181**(23), 7221-7.
- Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312**(5778), 1355-9.
- Gordon, C. L., and King, J. (1994). Genetic properties of temperature-sensitive folding mutants of the coat protein of phage P22. *Genetics* **136**(2), 427-38.
- Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**(14), 4121-30.
- Hatfull, G. F., and Sarkis, G. J. (1993). DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol* **7**(3), 395-405.
- Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E., and Liljas, L. (2003). The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J Mol Biol* **334**(5), 885-99.
- Henderson, P. J., and Giddens, R. A. (1977). 2-Deoxy-D-galactose, a substrate for the galactose-transport system of *Escherichia coli*. *Biochem J* **168**(1), 15-22.
- Henderson, P. J., Giddens, R. A., and Jones-Mortimer, M. C. (1977). Transport of galactose, glucose and their molecular analogues by *Escherichia coli* K12. *Biochem J* **162**(2), 309-20.

- Hendrix, R. W., and Duda, R. L. (1998). Bacteriophage HK97 head assembly: a protein ballet. *Adv Virus Res* **50**, 235-88.
- Ishii, T., and Yanagida, M. (1975). Molecular organization of the shell of the T-even bacteriophage head. *J Mol Biol* **97**(4), 655-60.
- Ishii, T., and Yanagida, M. (1977). The two dispensable structural proteins (soc and hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro. *J Mol Biol* **109**(4), 487-514.
- Iwasaki, K., Trus, B. L., Wingfield, P. T., Cheng, N., Campusano, G., Rao, V. B., and Steven, A. C. (2000). Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein, Soc. *Virology* **271**(2), 321-33.
- Jiang, W., Li, Z., Zhang, Z., Baker, M. L., Prevelige, P. E., Jr., and Chiu, W. (2003). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol* **10**(2), 131-5.
- Johnson, J. E., and Chiu, W. (2007). DNA packaging and delivery machines in tailed bacteriophages. *Curr Opin Struct Biol* **17**(2), 237-43.
- Juhala, R. J., Ford, M. E., Duda, R. L., Youton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299**(1), 27-51.
- Kang, H. J., and Baker, E. N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *J Biol Chem* **284**(31), 20729-37.
- Kang, H. J., Coulibaly, F., Clow, F., Proft, T., and Baker, E. N. (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* **318**(5856), 1625-8.
- Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H., and Baker, E. N. (2009). The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci U S A* **106**(40), 16967-71.
- Karakousis, G., Ye, N., Li, Z., Chiu, S. K., Reddy, G., and Radding, C. M. (1998). The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. *J Mol Biol* **276**(4), 721-31.
- Kropinski, A. M. (2000). Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J Bacteriol* **182**(21), 6066-74.
- Lander, G. C., Evilevitch, A., Jeembaeva, M., Potter, C. S., Carragher, B., and Johnson, J. E. (2008). Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. *Structure* **16**(9), 1399-406.
- Lata, R., Conway, J. F., Cheng, N., Duda, R. L., Hendrix, R. W., Wikoff, W. R., Johnson, J. E., Tsuruta, H., and Steven, A. C. (2000). Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell* **100**(2), 253-63.
- Lee, K. K., Gan, L., Tsuruta, H., Hendrix, R. W., Duda, R. L., and Johnson, J. E. (2004). Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid. *J Mol Biol* **340**(3), 419-33.
- Lee, K. K., Gan, L., Tsuruta, H., Moyer, C., Conway, J. F., Duda, R. L., Hendrix, R. W., Steven, A. C., and Johnson, J. E. (2008). Virus capsid expansion driven by the capture of mobile surface loops. *Structure* **16**(10), 1491-502.
- Lee, K. N., Lee, C. S., Tae, W. C., Jackson, K. W., Christiansen, V. J., and McKee, P. A. (2001). Crosslinking of alpha 2-antiplasmin to fibrin. *Ann N Y Acad Sci* **936**, 335-9.

- Lee, S. C., Koh, H., and Yu, M. H. (1991). Molecular properties of global suppressors of temperature-sensitive folding mutations in P22 tailspike endorhamnosidase. *J Biol Chem* **266**(34), 23191-6.
- Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2003). Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**(11), 2356-70.
- Li, Y., Conway, J. F., Cheng, N., Steven, A. C., Hendrix, R. W., and Duda, R. L. (2005). Control of virus assembly: HK97 "Whiffleball" mutant capsids without pentons. *J Mol Biol* **348**(1), 167-82.
- Li, Z., Karakousis, G., Chiu, S. K., Reddy, G., and Radding, C. M. (1998). The beta protein of phage lambda promotes strand exchange. *J Mol Biol* **276**(4), 733-44.
- Liu, P., Jenkins, N. A., and Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**(3), 476-84.
- Lokhandwala, P. M., Nguyen, T. L., Bowzard, J. B., and Craven, R. C. (2008). Cooperative role of the MHR and the CA dimerization helix in the maturation of the functional retrovirus capsid. *Virology* **376**(1), 191-8.
- Lorand, L., and Graham, R. M. (2003). Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nat Rev Mol Cell Biol* **4**(2), 140-56.
- Maurides, P. A., Schwarz, J. J., and Berget, P. B. (1990). Intragenic suppression of a capsid assembly-defective P22 tailspike mutation. *Genetics* **125**(4), 673-81.
- Mitraki, A., Danner, M., King, J., and Seckler, R. (1993). Temperature-sensitive mutations and second-site suppressor substitutions affect folding of the P22 tailspike protein in vitro. *J Biol Chem* **268**(27), 20071-5.
- Morais, M. C., Choi, K. H., Koti, J. S., Chipman, P. R., Anderson, D. L., and Rossmann, M. G. (2005). Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell* **18**(2), 149-59.
- Mosesson, M. W., Siebenlist, K. R., Hainfeld, J. F., and Wall, J. S. (1995). The covalent structure of factor XIIIa crosslinked fibrinogen fibrils. *J Struct Biol* **115**(1), 88-101.
- Muniyappa, K., and Radding, C. M. (1986). The homologous recombination system of phage lambda. Pairing activities of beta protein. *J Biol Chem* **261**(16), 7472-8.
- Murialdo, H., and Tzamtzis, D. (1997). Mutations of the coat protein gene of bacteriophage lambda that overcome the necessity for the FI gene; the EFi domain. *Mol Microbiol* **24**(2), 341-53.
- Murialdo, H., Tzamtzis, D., Berru, M., Fife, W. L., and Becker, A. (1997). Mutations in the terminase genes of bacteriophage lambda that bypass the necessity for FI. *Mol Microbiol* **24**(5), 937-52.
- Murthy, S. N., Wilson, J., Guy, S. L., and Lorand, L. (1991). Intramolecular crosslinking of monomeric fibrinogen by tissue transglutaminase. *Proc Natl Acad Sci U S A* **88**(23), 10601-4.
- Murthy, S. N., Wilson, J. H., Lukas, T. J., Veklich, Y., Weisel, J. W., and Lorand, L. (2000). Transglutaminase-catalyzed crosslinking of the Aalpha and gamma constituent chains in fibrinogen. *Proc Natl Acad Sci U S A* **97**(1), 44-8.
- Nagano, N., Ota, M., and Nishikawa, K. (1999). Strong hydrophobic nature of cysteine residues in proteins. *FEBS Lett* **458**(1), 69-71.
- Navas-Martin, S., Hingley, S. T., and Weiss, S. R. (2005). Murine coronavirus evolution in vivo: functional compensation of a detrimental amino acid substitution in the receptor binding domain of the spike glycoprotein. *J Virol* **79**(12), 7629-40.

- Oppenheim, A. B., Rattray, A. J., Bubunencko, M., Thomason, L. C., and Court, D. L. (2004). In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* **319**(2), 185-9.
- Parent, K. N., Ranaghan, M. J., and Teschke, C. M. (2004). A second-site suppressor of a folding defect functions via interactions with a chaperone network to improve folding and assembly in vivo. *Mol Microbiol* **54**(4), 1036-50.
- Parent, K. N., Suhanovsky, M. M., and Teschke, C. M. (2007). Polyhead formation in phage P22 pinpoints a region in coat protein required for conformational switching. *Mol Microbiol* **65**(5), 1300-10.
- Petrov, A. S., and Harvey, S. C. (2008). Packaging double-helical DNA into viral capsids: structures, forces, and energetics. *Biophys J* **95**(2), 497-502.
- Platt, T. (1984). Toxicity of 2-deoxygalactose to *Saccharomyces cerevisiae* cells constitutively synthesizing galactose-metabolizing enzymes. *Mol Cell Biol* **4**(5), 994-6.
- Popa, M. P., McKelvey, T. A., Hempel, J., and Hendrix, R. W. (1991a). Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. *J Virol* **65**(6), 3227-37.
- Poteete, A. R., and Fenton, A. C. (1993). Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. *Genetics* **134**(4), 1013-21.
- Poteete, A. R., Fenton, A. C., and Murphy, K. C. (1988). Modulation of *Escherichia coli* RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. *J Bacteriol* **170**(5), 2012-21.
- Purdy, J. G., Flanagan, J. M., Ropson, I. J., Rennoll-Bankert, K. E., and Craven, R. C. (2008). Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly. *J Virol* **82**(12), 5951-61.
- Qin, L., Fokine, A., O'Donnell, E., Rao, V. B., and Rossmann, M. G. (2009). Structure of the Small Outer Capsid Protein, Soc: A Clamp for Stabilizing Capsids of T4-like Phages. *J Mol Biol*.
- Radding, C. M. (1973). Molecular mechanisms in genetic recombination. *Annu Rev Genet* **7**, 87-111.
- Rader, A. J., Vlad, D. H., and Bahar, I. (2005). Maturation dynamics of bacteriophage HK97 capsid. *Structure* **13**(3), 413-21.
- Rao, V. B., and Feiss, M. (2008). The bacteriophage DNA packaging motor. *Annu Rev Genet* **42**, 647-81.
- Ross, P. D., Cheng, N., Conway, J. F., Firek, B. A., Hendrix, R. W., Duda, R. L., and Steven, A. C. (2005). Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. *Embo J* **24**(7), 1352-63.
- Rossmann, M. G., Arisaka, F., Battisti, A. J., Bowman, V. D., Chipman, P. R., Fokine, A., Hafenstein, S., Kanamaru, S., Kostyuchenko, V. A., Mesyanzhinov, V. V., Shneider, M. M., Morais, M. C., Leiman, P. G., Palermo, L. M., Parrish, C. R., and Xiao, C. (2007). From structure of the complex to understanding of the biology. *Acta Crystallogr D Biol Crystallogr* **63**(Pt 1), 9-16.
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y. H., Falcon, L. I., Souza, V.,

- Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Neilson, K., Friedman, R., Frazier, M., and Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**(3), e77.
- Russel, M. (1993). Protein-protein interactions during filamentous phage assembly. *J Mol Biol* **231**(3), 689-97.
- Schoepp, R. J., and Johnston, R. E. (1993). Sindbis virus pathogenesis: phenotypic reversion of an attenuated strain to virulence by second-site intragenic suppressor mutations. *J Gen Virol* **74** (Pt 8), 1691-5.
- Sternberg, N., and Weisberg, R. (1977). Packaging of coliphage lambda DNA. II. The role of the gene D protein. *J Mol Biol* **117**(3), 733-59.
- Steven, A. C., Greenstone, H. L., Booy, F. P., Black, L. W., and Ross, P. D. (1992). Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, co-operativity, and super-stabilization by soc binding. *J Mol Biol* **228**(3), 870-84.
- Tatsukawa, H., Fukaya, Y., Frampton, G., Martinez-Fuentes, A., Suzuki, K., Kuo, T. F., Nagatsuma, K., Shimokado, K., Okuno, M., Wu, J., Iismaa, S., Matsuura, T., Tsukamoto, H., Zern, M. A., Graham, R. M., and Kojima, S. (2009). Role of transglutaminase 2 in liver injury via cross-linking and silencing of transcription factor Sp1. *Gastroenterology* **136**(5), 1783-95 e10.
- Thirion, J. P., Banville, D., and Noel, H. (1976). Galactokinase mutants of Chinese hamster somatic cells resistant to 2-deoxygalactose. *Genetics* **83**(1), 137-147.
- Thomason, L., Court, D. L., Bubunenko, M., Costantino, N., Wilson, H., Datta, S., and Oppenheim, A. (2007). Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol* **Chapter 1**, Unit 1 16.
- Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. *Trends Microbiol* **12**(5), 228-34.
- Villafane, R., Fleming, A., and Haase-Pettingell, C. (1994). Isolation of suppressors of temperature-sensitive folding mutations. *J Bacteriol* **176**(1), 137-42.
- Vlaycheva, L., Nickells, M., Droll, D. A., and Chambers, T. J. (2004). Yellow fever 17D virus: pseudo-revertant suppression of defective virus penetration and spread by mutations in domains II and III of the E protein. *Virology* **327**(1), 41-9.
- Wackernagel, W., and Radding, C. M. (1973). Transfection by half molecules and inverted molecules of lambda DNA: requirement for exo and -promoted recombination. *Virology* **52**(2), 425-32.
- Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**(4), e36.
- Wikoff, W. R., Che, Z., Duda, R. L., Hendrix, R. W., and Johnson, J. E. (2003). Crystallization and preliminary analysis of a dsDNA bacteriophage capsid intermediate: Prohead II of HK97. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 12), 2060-4.
- Wikoff, W. R., Conway, J. F., Tang, J., Lee, K. K., Gan, L., Cheng, N., Duda, R. L., Hendrix, R. W., Steven, A. C., and Johnson, J. E. (2006). Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. *J Struct Biol* **153**(3), 300-6.



- Wikoff, W. R., Duda, R. L., Hendrix, R. W., and Johnson, J. E. (1998). Crystallization and preliminary X-ray analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Virology* **243**(1), 113-8.
- Wikoff, W. R., Duda, R. L., Hendrix, R. W., and Johnson, J. E. (1999). Crystallographic analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 4), 763-71.
- Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**(5487), 2129-33.
- Wong, Q. N., Ng, V. C., Lin, M. C., Kung, H. F., Chan, D., and Huang, J. D. (2005). Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic Acids Res* **33**(6), e59.
- Xie, Z., and Hendrix, R. W. (1995). Assembly in vitro of bacteriophage HK97 proheads. *J Mol Biol* **253**(1), 74-85.
- Yang, Q., Maluf, N. K., and Catalano, C. E. (2008). Packaging of a unit-length viral genome: the role of nucleotides and the gpD decoration protein in stable nucleocapsid assembly in bacteriophage lambda. *J Mol Biol* **383**(5), 1037-48.
- Yao, J., and Gillam, S. (2000). A single-amino-acid substitution of a tyrosine residue in the rubella virus E1 cytoplasmic domain blocks virus release. *J Virol* **74**(7), 3029-36.
- Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**(11), 5978-83.
- Yu, D., Sawitzke, J. A., Ellis, H., and Court, D. L. (2003). Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci U S A* **100**(12), 7207-12.
- Yuzenkova, J., Nechaev, S., Berlin, J., Rogulja, D., Kuznedelov, K., Inman, R., Mushegian, A., and Severinov, K. (2003). Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J Mol Biol* **330**(4), 735-48.
- Zhang, P., Li, D., Stewart-Jones, G., Shao, X., Zhang, Y., Chen, Q., Li, Y., He, Y. W., Xu, X. N., and Zhang, H. T. (2009). A single amino acid defines cross-species reactivity of tree shrew (*Tupaia belangeri*) CD1d to human invariant natural killer T (iNKT) cells. *Immunology* **128**(4), 500-10.
- Zhang, T., Breitbart, M., Lee, W. H., Run, J. Q., Wei, C. L., Soh, S. W., Hibberd, M. L., Liu, E. T., Rohwer, F., and Ruan, Y. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol* **4**(1), e3.
- Li, Y. (2000) Genetic and functional analysis of the major capsid protein of bacteriophage HK97. PhD Thesis
- Baros, B. (2002) Mutational analysis of bacteriophage HK97 capsid structure and assembly. PhD Thesis
- Casjens, S. and Hendrix, R. (1988) Control mechanisms in dsDNA bacteriophages assembly. *The Bacteriophages*, Vol. 1